

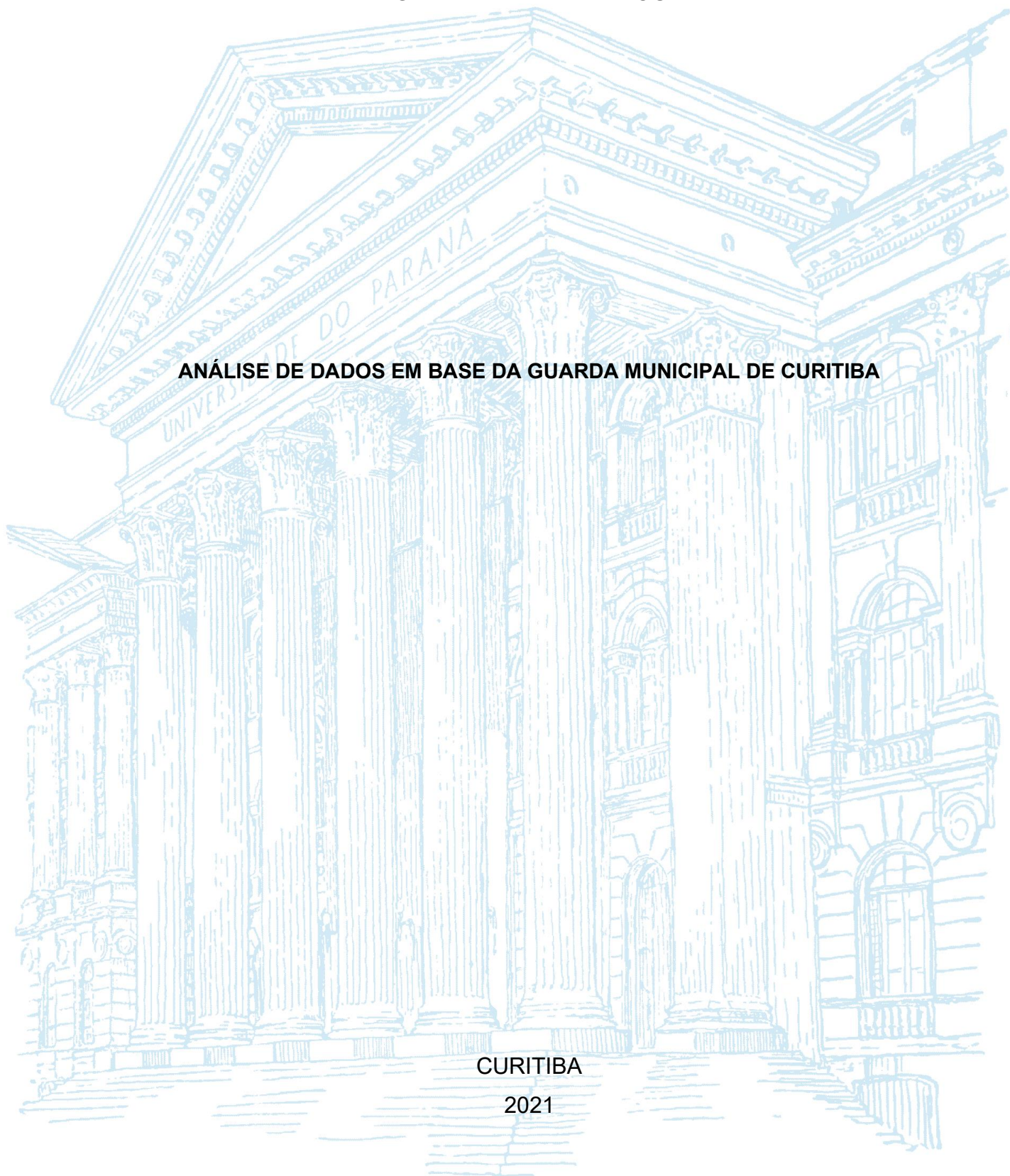
UNIVERSIDADE FEDERAL DO PARANÁ

ANGELA FERREIRA DE SOUZA

**ANÁLISE DE DADOS EM BASE DA GUARDA MUNICIPAL DE CURITIBA**

CURITIBA

2021



ANGELA FERREIRA DE SOUZA

**ANÁLISE DE DADOS EM BASE DA GUARDA MUNICIPAL DE CURITIBA**

Trabalho de Conclusão de Curso apresentado ao curso de Gestão da Informação, do Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Gestão da Informação.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Denise Fukumi Tsunoda

CURITIBA  
2021

## **AGRADECIMENTOS**

Primeiramente eu gostaria de agradecer a Deus por ter iluminado o meu caminho sempre, me dando saúde e forças para superar todos os momentos difíceis ao longo da minha graduação e o atual momento de pandemia do Covid-19.

Sou grata pela confiança depositada pela minha orientadora Prof.<sup>a</sup> Dr.<sup>a</sup> Denise Fukumi Tsunoda pela sua dedicação em inúmeras horas, me apoiando e aconselhando durante o projeto para sanar as minhas questões e me colocar na direção correta.

Agradeço aos meus pais e minha irmã por sempre estarem ao meu lado, acreditando em mim que eu seria capaz de superar os obstáculos e me apoiar nas minhas conquistas.

Ao meu namorado que sempre esteve ao meu lado, e nunca deixou que eu desistisse dos meus sonhos e me apoiou muito durante o meu percurso acadêmico.

Ao meu sogro, sogra, cunhado e cunhada por sempre me apoiarem e estarem ao meu lado.

Agradeço também ao apoio que recebi de todos no meu trabalho, ao mestre em estatística e o Mestre em Medicina do meu trabalho por me aconselharem e me apoiarem nesta jornada.

Por fim aos meus amigos que sempre estiveram presentes do início ao fim e a Universidade Federal do Paraná com e todo o seu corpo docente e funcionários.

## RESUMO

O presente estudo utiliza técnicas de análise de dados na base de dados da Guarda Municipal de Curitiba por meio de estatística descritiva e mineração de dados, com ocorrências atendidas no período de 2009 até 2019. Foi utilizado o KDD (Knowledge-Discovery in Databases) com os algoritmos Rpart e Naïve Bayes para análise da base de dados. Para a aplicação dos algoritmos foi necessário realizar agrupamentos nas descrições das ocorrências atendidas resultando em duas bases de dados, uma base de dados agrupada manualmente e uma base de dados agrupada pelo K-Means. A base de dados agrupada pelo K-Means obteve uma taxa de acerto de 62% em ambos os algoritmos, superior aos resultados com a base de dados agrupada manualmente que obteve 46% na árvore de decisão e 48% no Naïve Bayes. O principal objetivo foi identificar tendências e padrões na base de dados. Como resultados, observou-se que sábado e domingo são os dias da semana que demandam maior atendimento às ocorrências; que a regional Matriz contém a maior parte das ocorrências (81.038); e foi identificada a média relativa de (21.791) ocorrências atendidas por mês. Finaliza com as contribuições como suporte à tomada de decisão na gestão urbana e indicativo para trabalhos futuros como aplicação de mapas interativos.

Palavras-chave: Análise de Dados. Guarda Municipal de Curitiba. Mineração de dados. Árvore de decisão. Naïve Bayes. KDD.



## **ABSTRACT**

The present study uses data analysis techniques in the database of the Municipal Guard of Curitiba through descriptive statistics and mining and data, with occurrences attended in the period from 2009 to 2019. This work adopts the KDD (Knowledge - Discovery in Databases) with the Rpart and Naïve Bayes algorithms for database analysis. For the application of the algorithms, it was necessary to perform groupings in the descriptions of the attended cases resulting in two databases, a database grouped manually and a database grouped by K-Means. The database grouped by K-Means obtained a 62% hit rate in both algorithms, higher than the results with the manually grouped database which obtained 46% in the decision tree and 48% in Naïve Bayes. The main objective was to identify trends and patterns in the database. As a result, it was observed that Saturday and Sunday are the days of the week that demand greater attendance to events; that the regional headquarters contains most of the occurrences (81,038); and the relative average of (21,791) occurrences attended per month was identified. It ends with contributions to support decision-making in urban management and indicative for future work such as the application of interactive maps.

**Keywords:** Data Analysis. Curitiba Municipal Guard. Data mining. Decision tree. Naïve Bayes. KDD.

## LISTA DE FIGURAS

FIGURA 1 - MODELO DE FLUXO DE INFORMAÇÃO .....	25
FIGURA 2 - O PROCESSO DE GERENCIAMENTO DA INFORMAÇÃO .....	26
FIGURA 3 - CICLO DE GESTÃO DA INFORMAÇÃO .....	27
FIGURA 4 - CICLO DE GESTÃO DA INFORMAÇÃO .....	28
FIGURA 5 - NÍVEIS HIERÁRQUICOS ORGANIZACIONAL .....	30
FIGURA 6 - PROCESSO DECISÓRIO SOB A ÓTICA DA ORGANIZAÇÃO .....	34
FIGURA 7 - DECISÕES SÃO ESCOLHAS QUE PROCURAM RESOLVER PROBLEMAS E APROVEITAR OPORTUNIDADES .....	35
FIGURA 8 - PROCESSO KDD .....	36
FIGURA 9 - SUBFASES DO PRÉ-PROCESSAMENTO .....	38
FIGURA 10 - TIPOS DE ATRIBUTOS .....	38
FIGURA 11 - PRINCIPAIS PROBLEMAS COM DADOS .....	39
FIGURA 12 - ETAPAS DO PROCESSO DE PREPARAÇÃO DA BASE DE DADOS .....	40
FIGURA 13 - MULTIDISCIPLINARIDADE DA MINERAÇÃO DE DADOS .....	41
FIGURA 14 - TIPO DE VARIÁVEIS .....	50
FIGURA 15 - EXEMPLO ÁRVORE DE DECISÃO .....	51
FIGURA 16 - AGRUPAMENTO DO ATRIBUTO NATUREZA_DESCRÇÃO .....	64
FIGURA 17 - SCRIPT DE AGRUPAMENTO K-MEANS .....	65
FIGURA 18 - SCRIPT DE CARGA DE DADOS .....	85
FIGURA 19 - SCRIPT DE INSTALAÇÃO E EXECUÇÃO DE PACOTES DA ÁRVORE DE DECISÃO .....	86
FIGURA 20 - SCRIPT DO ALGORITMO DE ÁRVORE DE DECISÃO .....	87
FIGURA 21 - SCRIPT DE VISUALIZAÇÃO DA ÁRVORE .....	87
FIGURA 22 - RESULTADO DA ÁRVORE NO CONSOLE DA BASE DE DADOS MANUAL .....	88
FIGURA 23 - RESULTADO DA ÁRVORE NO CONSOLE DA BASE DE DADOS K-MEANS .....	88
FIGURA 24 - ÁRVORE DE DECISÃO COM DADOS AGRUPADOS MANUAL .....	89
FIGURA 25 - ÁRVORE DE DECISÃO COM DADOS AGRUPADOS K-MEANS .....	91
FIGURA 26 - SCRIPT DE GERAÇÃO DA MATRIZ CONFUSÃO DA ÁRVORE DE DECISÃO .....	92
FIGURA 27 - MATRIZ CONFUSÃO DA ÁRVORE DOS DADOS AGRUPADOS MANUAL .....	92
FIGURA 28 - MATRIZ CONFUSÃO DA ÁRVORE DOS DADOS AGRUPADOS KMEANS .....	93
FIGURA 29 - SCRIPT DE INSTALAÇÃO E EXECUÇÃO DE PACOTES DO NAÏVE BAYES .....	94
FIGURA 30 - SCRIPT DO ALGORITMO DE NAÏVE BAYES .....	94
FIGURA 31 - DISTRIBUIÇÃO DE PROBABILIDADES POR GRUPO NA BASE MANUAL DO NAÏVE BAYES .....	95
FIGURA 32 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ATENDIMENTO_ANO NA BASE MANUAL .....	95
FIGURA 33 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRÊNCIA_MES NA BASE MANUAL .....	96
FIGURA 34 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA_DIA_SEMANA NA BASE MANUAL .....	96

FIGURA 35 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA_HORA NA BASE MANUAL .....	96
FIGURA 36 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO REGIONAL_FATO_NOME NA BASE MANUAL .....	97
FIGURA 37 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO FLAG_FLAGRANTE NA BASE MANUAL .....	97
FIGURA 38 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO NATUREZA_DEFESA_CIVIL NA BASE MANUAL .....	98
FIGURA 39 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ORIGEM_CHAMADO_DESCRICAO NA BASE MANUAL .....	98
FIGURA 40 - SCRIPT DE GERAÇÃO DA MATRIZ CONFUSÃO DO NAÏVE BAYES .....	99
FIGURA 41 - MATRIZ CONFUSÃO DO NAÏVE BAYES DOS DADOS AGRUPADOS MANUAL.....	99
FIGURA 42 - DISTRIBUIÇÃO DE PROBABILIDADES POR GRUPO NA BASE K-MEANS .....	100
FIGURA 43 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ATENDIMENTO_ANO NA BASE K-MEANS .....	100
FIGURA 44 DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA_MES NA BASE K-MEANS .....	100
FIGURA 45 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA_DIA_SEMANA NA BASE K-MEANS .....	101
FIGURA 46 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA_HORA NA BASE K-MEANS.....	101
FIGURA 47 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO REGIONAL_FATO_NOME NA BASE K-MEANS.....	102
FIGURA 48 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO FLAG_FLAGRANTE NA BASE K-MEANS .....	102
FIGURA 49 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO NATUREZA_DEFESA_CIVIL NA BASE K-MEANS .....	103
FIGURA 50 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ORIGEM_CHAMADO_DESCRICAO NA BASE K .....	103
FIGURA 51 - MATRIZ CONFUSÃO DO NAÏVE BAYES DOS DADOS AGRUPADOS KMEANS .....	104

## LISTA DE QUADROS

QUADRO 1 - ATUAÇÃO DA GUARDA MUNICIPAL E O EXERCÍCIO DO PODER DA POLÍCIA .....	20
QUADRO 2 - DIFERENÇA ENTRE DADOS, INFORMAÇÃO E CONHECIMENTO .....	24
QUADRO 3 - MODELO RELACIONAL E COMPORTAMENTAL DE TOMADA DE DECISÃO .....	31
QUADRO 4 - PRINCIPAIS TIPOS DE DECISÕES .....	33
QUADRO 5 - TÉCNICAS DE MINERAÇÃO DE DADOS .....	42
QUADRO 6 - HEURÍSTICAS DA TÉCNICA DESCRITIVA .....	43
QUADRO 7 - HEURÍSTICAS DA TÉCNICA PREDITIVA .....	44
QUADRO 8 - CLASSIFICAÇÃO DAS PESQUISAS .....	46
QUADRO 9 - QUADRO DE ATRIBUTOS DA BASE DE DADOS DA GUARDA MUNICIPAL DE CURITIBA .....	47
QUADRO 10 - RELAÇÃO DE VALIDADORES DA BASE DE DADOS .....	49
QUADRO 11 - ATRIBUTOS SELECIONADOS PARA COMPOR A BASE DE DADOS .....	52
QUADRO 12 - ATRIBUTOS RETIRADOS DA ANÁLISE DOS DADOS .....	53
QUADRO 13 - RELAÇÃO DE BAIRROS NAS REGIONAIS 1 .....	57
QUADRO 14 - RELAÇÃO DE BAIRROS NAS REGIONAIS 2 .....	58
QUADRO 15 - INSTÂNCIAS NOMINAIS NORMALIZADAS .....	61
QUADRO 16 - CATEGORIZAÇÃO DE VARIÁVEIS NO ATRIBUTO ORIGEM_CHAMADO_DESCRICAO .....	62
QUADRO 17 - DISTRIBUIÇÃO DE HORAS POR PERÍODO .....	63
QUADRO 18 - CATEGORIZAÇÃO DO ATRIBUTO OCORRÊNCIA MÊS .....	63
QUADRO 19 - CLASSIFICAÇÃO DOS ATRIBUTOS .....	66

## LISTA DE TABELAS

TABELA 1 - PESQUISA EM BIBLIOTECAS DE TESES E DISSERTAÇÕES .....	17
TABELA 2 - SELEÇÃO DOS PRINCIPAIS RESULTADOS RELACIONADO AO TEMA NAS BIBLIOTECAS .....	17
TABELA 3 - RELAÇÃO DE VALORES AUSENTES NO ATRIBUTO NATUREZA DESCRIÇÃO E SUBCATEGORIA DESCRIÇÃO .....	55
TABELA 4 - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL TATUQUARA .....	56
TABELA 5 - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BAIRRO NOVO .....	59
TABELA 6 - RELAÇÃO DE INCONSISTÊNCIAS ENTRE AS REGIONAIS .....	59
TABELA 7 - RELAÇÃO DE DADOS ELIMINADOS APLICADOS EM QUATRO VALIDAÇÕES .....	61
TABELA 8 - NORMALIZAÇÃO DAS INSTANCIAS DO ATRIBUTO NATUREZA1_DEFESA_CIVIL .....	62
TABELA 9 - CLASSIFICAÇÃO DOS GRUPOS BASE MANUAL .....	64
TABELA 10 - CLASSIFICAÇÃO DOS GRUPELO PELO K-MEANS .....	66
TABELA 11 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE ORIGENS DE OCORRÊNCIAS POR ANO .....	68
TABELA 12 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE SECRETARIAS/ANO .....	68
TABELA 13 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE FLAGRANTES POR ANOS .....	69
TABELA 14 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE DEFESA CIVIL POR ANOS .....	70
TABELA 15 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE DESCRIÇÕES POR ANOS .....	72
TABELA 16 - REDISTRIBUIÇÃO DA CLASSIFICAÇÃO DOS GRUPELO PELO MANUAL .	73
TABELA 17 - REDISTRIBUIÇÃO DA CLASSIFICAÇÃO DOS GRUPOS PELO K-MEANS .	73
TABELA 18 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE SUBCATEGORIAS POR ANOS .....	75
TABELA 19 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE OCORRENCIA_DIA_SEMANA POR ANOS .....	76
TABELA 20 - FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE OCORRÊNCIA HORA POR ANOS .....	78
TABELA 21 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE REGIONAL POR ANOS .....	79
TABELA 22 - FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DA REGIONAL MATRIZ POR BAIRROS .....	80
TABELA 23 - TABELA FREQUÊNCIAS ABSOLUTAS DA DISTRIBUIÇÃO DE BAIRROS POR ANO .....	82
TABELA 24 - FREQUÊNCIA ABSOLUTAS DAS 10PRINCIPAIS DESCRIÇÕES POR SUBCATEGORIA .....	83
TABELA 25 - DISTRIBUIÇÃO DE VARIAÇÃO, MÍNIMO E MÁXIMO DE OCORRÊNCIAS..	84

## LISTA DE GRÁFICOS

GRÁFICO 1 - GRÁFICO DE DISTRIBUIÇÃO DE OCORRÊNCIAS POR ANO .....	67
GRÁFICO 2 - GRÁFICO DA DISTRIBUIÇÃO DAS 10 PRINCIPAIS DESCRIÇÕES POR ANO .....	71
GRÁFICO 3 - GRÁFICO DE DISTRIBUIÇÃO DAS 10 PRINCIPAIS SUBCATEGORIAS1_DESCRIÇÃO .....	74
GRÁFICO 4 - DISTRIBUIÇÃO DAS 10 PRINCIPAIS OCORRENCIA_MES.....	75
GRÁFICO 5 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR DIA DA SEMANA NOS ANOS ...	77
GRÁFICO 6 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR HORA NOS ANOS.....	78
GRÁFICO 7 - GRÁFICO DE DISTRIBUIÇÃO DE OCORRÊNCIAS POR REGIONAL NOS ANOS .....	79
GRÁFICO 8 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR BAIRROS NOS ANOS .....	81



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	PROBLEMATIZAÇÃO	15
1.2	OBJETIVOS	16
1.3	JUSTIFICATIVA	16
1.4	DELIMITAÇÃO DA PESQUISA	18
1.5	ESTRUTURA DO DOCUMENTO	19
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>20</b>
2.1	GUARDA MUNICIPAL	20
2.1.1	Guarda Municipal de Curitiba	21
2.2	LEI DE ACESSO À INFORMAÇÃO	22
2.3	GESTÃO DA INFORMAÇÃO	23
2.3.1	Recuperação da informação	29
2.3.2	Tomada de decisão	31
2.4	KDD ( <i>KNOWLEDGE DISCOVERY IN DATABASE</i> )	36
2.4.1	Pré-processamento	37
2.4.2	Mineração de dados	41
2.4.3	Pós-processamento	44
<b>3</b>	<b>METODOLOGIA</b>	<b>45</b>
3.1	CARACTERIZAÇÃO DA PESQUISA	45
3.2	AMBIENTE DE PESQUISA	46
3.3	DICIONÁRIO DE DADOS	46
3.4	PRÉ PROCESSAMENTO	48
3.4.1	Limpeza dos dados	48
3.4.2	Tratamento da base de dados	49
3.5	ESTATÍSTICA DESCRITIVA	50
3.6	PÓS PROCESSAMENTO	51
<b>4.</b>	<b>EXPERIMENTOS E ANÁLISE DOS RESULTADO</b>	<b>52</b>
4.1	PRÉ-PROCESSAMENTO	52
4.1.1	Limpeza dos dados	54
4.1.2	Tratamento da base de dados	61
4.2	ESTATÍSTICA DESCRITIVA	66
4.3	MINERAÇÃO DE DADOS	85
4.3.1	Árvore de Decisão	86
4.3.2	Naïve Bayes	93

4.4	PÓS PROCESSAMENTO	104
4.4.1	Resultados do algoritmo árvore de decisão	106
4.4.2	Resultados do algoritmo Naïve Bayes	107
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>108</b>
5.1	ALCANCE DOS OBJETIVOS	108
5.2	CONTRIBUIÇÕES	110
5.3	TRABALHOS FUTUROS	111
	<b>REFERÊNCIAS</b>	<b>112</b>
	<b>APÊNDICE</b>	<b>117</b>
	<b>APÊNDICE A - ATRIBUTOS E INSTÂNCIAS DA BASE DE DADOS DAS OCORRÊNCIAS DA GUARDA MUNICIPAL DE CURITIBA</b>	<b>117</b>
	<b>APÊNDICE B - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BOA VISTA</b>	<b>124</b>
	<b>APÊNDICE C - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BOQUEIRÃO</b>	<b>125</b>
	<b>APÊNDICE D - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL CAJURU</b>	<b>125</b>
	<b>APÊNDICE E - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL CIC</b>	<b>126</b>
	<b>APÊNDICE F - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL MATRIZ</b>	<b>126</b>
	<b>APÊNDICE G - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL PORTÃO</b>	<b>127</b>
	<b>APÊNDICE H - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL PINHEIRINHO</b>	<b>128</b>
	<b>APÊNDICE I - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL SANTA FELICIDADE</b>	<b>129</b>
	<b>APÊNDICE J- ANÁLISE DE INCONSISTÊNCIA DA REGIÃO METROPOLITANA</b>	<b>130</b>
	<b>APÊNDICE K - AGRUPAMENTO MANUAL DAS DESCRIÇÕES</b>	<b>132</b>
	<b>APÊNDICE L - AGRUPAMENTO DO K-MEANS DAS DESCRIÇÕES</b>	<b>135</b>
	<b>APÊNDICE M - RELAÇÃO DE ORIGEM DAS OCORRÊNCIAS POR ANOS</b>	<b>140</b>
	<b>APÊNDICE N- RELAÇÃO DE OCORRÊNCIAS POR SECRETARIA EM ANOS</b>	<b>141</b>
	<b>APÊNDICE O - RELAÇÃO DE DESCRIÇÕES POR ANOS</b>	<b>142</b>
	<b>APÊNDICE P - RELAÇÃO DE BAIRROS POR ANOS</b>	<b>146</b>
	<b>APÊNDICE Q - RELAÇÃO DA DESCRIÇÃO COM SUBCATEGORIA</b>	<b>149</b>
	<b>APÊNDICE R – RELAÇÃO SUBCATEGORIAS POR ANOS</b>	<b>151</b>
	<b>APÊNDICE S – SCRIPT DO R STUDIO AGRUPAMENTO DA BASE PELO K-MEANS</b>	<b>158</b>
	<b>APÊNDICE T – SCRIPT DO R STUDIO ALGORITMO ÁRVORE DE DECISÃO</b>	<b>159</b>
	<b>APÊNDICE U – SCRIPT DO R STUDIO ALGORITMO NAÏVE BAYES</b>	<b>160</b>

## 1 INTRODUÇÃO

O momento de pandemia do vírus “SARS-CoV-2” traz à lembrança de uma situação parecida que ocorreu entre 1918 a 1919 com a “gripe espanhola”. A diferença entre os dois casos citados é que no passado as informações eram registradas em documentos escritos e muitos deles foram perdidos ao longo da história (SAYÃO, 2012). Por meio das tecnologias digitais, os estudos ficaram muito mais avançados que os daquela época, além é claro, que através dos anos a evolução digital tem ganhado cada vez mais espaço, e ela é caracterizada pela intensificação no volume de informações, formatos e de tecnologias.

Quanto ao quesito da análise de dados especificamente, o caso brasileiro não difere muito do exemplo da gripe espanhola da área da saúde em relação a segurança. No Brasil a segurança dos dados sempre foi e é um tema muito discutido e relevante, porém o sentimento de insegurança permanece comum aos cidadãos. A análise de dados na área de segurança traz benefícios, inclusive por exemplo, por meio de predições de situações.

Como uma forma de minimizar esse sentimento de insegurança, a Guarda Municipal foi criada em 14 de junho de 1831 no Brasil com o primeiro objetivo de proteger o patrimônio público nas cidades (CARVALHO, 2011, p. 3) porém com os anos, cada vez mais os guardas foram ganhando responsabilidades sobre outros problemas enfrentados pela sociedade. No dia 17 de julho de 1986 foi elaborada e aprovada uma proposta para a implementação da Guarda Municipal de Curitiba, criada pelo prefeito da época Roberto Requião (CARVALHO, 2011, p. 13). Os guardas de Curitiba exerciam na época a vigilância interna e externa sobre patrimônios públicos, além de protegê-los do crime, orientar o público no trânsito, prevenir o vandalismo e fiscalizar os problemas na cidade.

Desde a implementação da Guarda Municipal em Curitiba foram atendidas inúmeras ocorrências, todas armazenadas em banco de dados. A partir de 2012, em decorrência da implantação da lei de acesso à informação, a base de dados de ocorrências da Guarda Municipal passou a ser livremente acessível aos cidadãos em geral.

De acordo com Soares, Jardim e Hermont (2013, p. 11) a Lei de Acesso à Informação (LAI) foi criada no ano de 2011 com o objetivo de tratar assuntos de

interesse da união, dos estados, dos distritos e municípios, disponibilizando informações do governo para a comunidade a fim de se realizar trocas de informações entre os cidadãos e garantindo uma maior participação de todos.

Dessa forma, partindo dessa preocupação com relação à segurança no país e utilizando uma base de dados de ocorrências atendidas da Guarda Municipal de Curitiba, a presente pesquisa objetiva analisar dados, identificar padrões e tendências em busca de oportunidades de aprimoramento no atendimento de ocorrências que chegam à Guarda Municipal de Curitiba.

## 1.1 PROBLEMATIZAÇÃO

Segundo Rezende (2002, p. 78), a sociedade do conhecimento tem as maiores fontes de riquezas em informação através da distribuição e manipulação da informação, que cresce de forma exponencial anualmente. Este volume crescente de informações promove a necessidade de se aplicar técnicas e métodos que os transformem em conhecimento, com vistas à tomada de decisão.

O contexto do autor traz que o governo e prefeituras geram a cada dia milhares de dados em diferentes áreas, e desde 2012 começaram a ser liberados muitos desses dados para os cidadãos em geral em decorrência da criação da lei de acesso à informação. Esta lei segundo Soares, Jardim, Hermont (2013, p. 11) disponibiliza dados de forma aberta com o objetivo de praticar transparências aos cidadãos, e a uma maior participação dos cidadãos nos processos, promovendo uma troca de informações entre o governo e o cidadão.

Porém, em relação ao interesse do indivíduo sobre o acesso aos dados abertos públicos, muitos cidadãos não têm o interesse sobre os dados tanto por falta de conhecimento, ou pela dificuldade em poder acessá-las, porque muitos órgãos do governo dificultam o acesso a estes dados, e com isso gera cada vez menos buscas pelas informações públicas.

De acordo com essa problemática, desde a implementação da Guarda Municipal em Curitiba, foram atendidas várias ocorrências, gerando um volume de dados. Com o elevado volume de dados surge a necessidade e oportunidade de realizar análises com o desígnio de identificar padrões e tendências. Nessa perspectiva torna-se importante responder à seguinte questão: **Como prever tendências nas ocorrências atendidas na Guarda Municipal de Curitiba?**

## 1.2 OBJETIVOS

A fim de se obter uma resposta para a questão de pesquisa apresentada acima, foram definidos objetivos gerais e específicos.

O objetivo geral é identificar tendências e padrões que possibilitem a descoberta de conhecimento de dados em uma base da Guarda Municipal de Curitiba.

O objetivo geral é desdobrado em objetivos específicos que são:

- a) realizar estudos com revisão sistemática de literatura;
- b) estudar a base e selecionar os principais atributos a serem analisados;
- c) calcular estatísticas descritivas para análise dos dados na base de dados;
- d) aplicar métodos de mineração de dados para descoberta de conhecimento;
- e) comparar e analisar os resultados obtidos por meio das análises estatísticas e mineração de dados.

## 1.3 JUSTIFICATIVA

De acordo com os objetivos deste trabalho apresentado no capítulo 1.2, a motivação da realização desta pesquisa é o anseio em contribuir na área da segurança, através de análise das bases de dados da Guarda Municipal de Curitiba.

A fim de se identificar trabalhos semelhantes foi realizado um levantamento em bases bibliográficas no dia 28 de dezembro de 2020 no portal da Universidade Federal do Paraná (EBSCOHOST), no portal da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e no portal *Scientific Electronic Library Online* (SciELO) de modo a verificar pesquisas existentes em relação ao tema. Em todas as pesquisas foi utilizado o filtro de região: Brasil. Para a realização da análise bibliográfica foram utilizadas as três bases bibliográficas e os resultados estão explicitados na Tabela 1.

TABELA 1 - PESQUISA EM BIBLIOTECAS DE TESES E DISSERTAÇÕES

Biblioteca	Termos	Tipo de Documento	Ano de Defesa	Resultados
EBSCOHOST	Todos os campos: "Guarda Municipal"	dissertação ou tese	2000-2020	97
CAPEL				217
SciELO				14
EBSCOHOST	Todos os campos: "Guarda Municipal" AND "Análise"	dissertação ou tese	2000-2020	2
CAPEL				65
SciELO				3
EBSCOHOST	Todos os campos: "Guarda Municipal" AND "Dados"	dissertação ou tese	2000-2020	4
CAPEL				64
SciELO				6

Fonte: Elaborado pela autora (2020)

Apesar dos valores terem sido considerados relevantes de acordo com a Tabela 1, ao gerar as buscas as bibliotecas trazem vários arquivos somente pelas palavras chaves, muitos destes arquivos não são relacionados ao tema portanto foi realizado uma análise selecionando todos os documentos relacionado ao tema conforme a Tabela 2.

TABELA 2 - SELEÇÃO DOS PRINCIPAIS RESULTADOS RELACIONADO AO TEMA NAS BIBLIOTECAS

Biblioteca	Termos	Resultados
EBSCOHOST	Todos os campos: "Guarda Municipal"	8
CAPEL		2
SciELO		1
EBSCOHOST	Todos os campos: "Guarda Municipal" AND "Análise"	1
CAPEL		
SciELO		1
EBSCOHOST	Todos os campos: "Guarda Municipal" AND "Dados"	2
CAPEL		3
SciELO		1

Fonte: Elaborado pela Autora (2020)



Na Tabela 2 é possível avaliar o baixo número de artigos (19 no total) de artigos relacionados ao tema. A maior parte não era relacionado ao assunto Guarda Municipal e os que discutiam não abordavam a análise dos dados.

Vivenciar o trabalho da Guarda Municipal em Curitiba (percebendo o cenário de atuação de atendimento de ocorrências) constituiu um dos fatores motivadores da realização desta pesquisa, a falta de artigos relacionado ao tema conforme analisado nas bibliotecas e o anseio em trabalhar com a área da segurança com análise dos dados complementa a motivação.

No âmbito da justificativa, para o curso de Gestão da Informação da UFPR, primeiramente a graduação destaca vários cursos relacionado com o tema deste trabalho como: Teoria da Informação, Mineração de dados, banco de dados, Métodos quantitativos de análise, Estatística, metodologia de pesquisa e entre outros. Este trabalho contempla várias partes adquiridas através destas disciplinas destacadas, contribuindo para um maior acervo de informações de estudo para a graduação de Gestão da Informação.

#### 1.4 DELIMITAÇÃO DA PESQUISA

A presente pesquisa abrangerá a identificação de padrões nas ocorrências atendidas pela Guarda Municipal de Curitiba, a pesquisa é de natureza qualitativa e tem como principal delimitação a descrição das ocorrências sendo categorizadas de forma manual e através do algoritmo K-Means.

A base de dados da guarda municipal de Curitiba tem as ocorrências atendidas com o período de 2009 até 2019. E a base apresenta o atributo Descrição e Subcategoria como principais descritores das ocorrências, e utiliza como atributo meta nos algoritmos selecionados.

Também é apresentada na base de dados a secretaria responsável pela ocorrência e a origem. Ainda é possível identificar a Regional, bairro, hora, dia da semana e ano de cada ocorrência registrada e as ocorrências estão colocadas se é de natureza civil e se foi pego em flagrante.

## 1.5 ESTRUTURA DO DOCUMENTO

O documento está estruturado em cinco seções principais. Nesta seção foram apresentadas a introdução, problema de pesquisa, objetivos (gerais e específicos), justificativa e delimitação da pesquisa.

A próxima apresenta a revisão de literatura e contempla os principais conceitos relacionados à pesquisa sendo: Guarda Municipal, Lei de Acesso à Informação (LAI), gestão da informação, recuperação da informação, tomada de decisão e os processos de descoberta de conhecimento em base de dados.

No capítulo 3 está presente a metodologia de pesquisa, que aborda a caracterização da pesquisa, ambiente de pesquisa, dicionário de dados e os processos do KDD sendo (pré-processamento, estatística, mineração de dados e pós processamento).

No capítulo 4 são apresentados os experimentos e análises dos resultados da base de dados. E por fim, as considerações finais presentes no capítulo 5 são divididas em três subseções sendo: alcance dos objetivos, contribuições e trabalhos futuros.

## 2 REVISÃO DA LITERATURA

Este capítulo apresenta os principais conceitos necessários para a compreensão da pesquisa realizada, a saber: guarda municipal, lei de acesso à informação, revisão de literatura sobre a gestão da informação, recuperação da informação, tomada de decisão e os processos de descoberta de conhecimento em bases de dados.

### 2.1 GUARDA MUNICIPAL

A Guarda Municipal segundo Carvalho (2011) possui, através da Constituição Federal de 1988, um poder municipal que corresponde à proteção ao patrimônio público. Ainda assim, cada município tem suas próprias legislações, dando diferentes poderes aos guardas municipais.

No Brasil, a Guarda Municipal está presente em quase todos os municípios. De Campos (2013, p. 60) extrai-se o Quadro 1, sobre as principais atuações da Guarda Municipal.

QUADRO 1 - ATUAÇÃO DA GUARDA MUNICIPAL E O EXERCÍCIO DO PODER DA POLÍCIA

Quanto a função	Quanto a atividade do poder da polícia	Quanto a característica da ação	Quanto ao objetivo da ação
Proteção aos bens, serviços e instalações.	Exercício da polícia administrativa	Preventiva e Repressiva	Garantir a municipalidade (as atividades inerentes às competências municipais).
Atividade Policial “estrito senso”	Exercício da polícia judiciária	Repressiva	Garantir a responsabilização dos violadores da ordem jurídica efetuando eventuais prisões em flagrante delito.

FONTE: Adaptado CAMPOS (2013, p.67)

De acordo com o Quadro 1 adaptado de Campos (2013, p. 67), o autor apresenta quatro tipos de atuação da Guarda Municipal:

- I. quanto a sua função, por exemplo relacionado a proteção de bens;
- II. diz respeito à atividade de poder da política e ela é o exército da polícia administrativa;
- III. corresponde à característica de ação e pode ser tanto preventiva e repressiva; e
- IV. quanto ao objetivo da ação por exemplo ela garante a municipalidade.

A constituição procurou integrar os municípios com a segurança pública de modo que é dever do estado, direito e responsabilidade de todos e que segundo o § 8o do art. 144: “Os Municípios poderão constituir guardas municipais destinadas à proteção de seus bens, serviços e instalações, conforme dispuser a lei”.

Campos (2013, p. 55) afirma que cada município tem sua certa autonomia responsável pela instituição das guardas municipais “conforme dispuser a lei”, e o autor ressalta que desnivelar as corporações “despadroniza” a Guarda Municipal de cada município. O autor ainda complementa que:

Os órgãos componentes do sistema de segurança pública são responsáveis pela manutenção da ordem, da segurança das pessoas e do seu patrimônio e deve ser exercido por órgãos policiais, ou seja, a Polícia Federal, Rodoviária e Ferroviária Federal e as Polícias Cíveis, Militares e o Corpo de Bombeiros. Como foi negada ao município a possibilidade de possuir esse instrumento legalmente constituído, há de concordar com a afirmação em voga que a Guarda Municipal “não é polícia”, logo se retira do mesmo a possibilidade de estas atuarem de forma completa nas duas vertentes previstas para os órgãos elencados no caput do art. 144 da CF/ 88. (Campos, 2013, p.55)

Ou seja, Campos (2013, p. 55) discorre que a Guarda Municipal não atua como polícia Federal, Rodoviária e Ferroviária, Cível, Militares e o Corpo de Bombeiros, mais sim mantém a manutenção da ordem, segurança das pessoas e patrimônios.

### 2.1.1 Guarda Municipal de Curitiba

A Guarda Municipal foi aprovada para implementação em Curitiba no dia 17 de julho de 1986 segundo Carvalho (2011, p. 143). Através do vereador José Maria Corrêa atendendo ao pedido do prefeito de Curitiba na época Roberto Requião foi implementada conforme a lei de número 6867/86 sob vários Artigos, com destaque as atribuições da Guarda Municipal:

Art. 2º São atribuições da Guarda Municipal:

I. exercer a vigilância interna e externa sobre os próprios municipais, parques, jardins, escolas, teatros, museus, bibliotecas, cemitérios, mercados, feiras-livres, no sentido de:

- a) protegê-los dos crimes contra o patrimônio;
- b) orientar o público e o trânsito de veículos, em caráter auxiliar à Polícia Militar;
- c) prevenir a ocorrência, internamente, de qualquer ilícito penal;
- d) controlar a entrada e a saída de veículos;
- e) prevenir sinistros, atos de vandalismo e danos ao patrimônio.

II. Garantir os serviços de responsabilidade do Município e, bem assim, sua ação fiscalizadora no desempenho de atividade de polícia administrativa, nos termos das Constituições Federal e Estadual e da Lei Orgânica. (Carvalho, 2011, p.143)

Carvalho (2011, p. 143) ressalta que o principal objetivo do projeto para a implementação da Guarda Municipal era minimizar as inseguranças em Curitiba e atender as necessidades da população seguindo as atribuições da Guarda Municipal citadas acima. Através da aprovação do projeto surgiu o VIGISERV (Serviço Municipal de vigilância) o qual com o passar dos anos se tornou a atual Guarda Municipal.

A Guarda Municipal desde sua implantação em Curitiba vem atendendo várias ocorrências todos os dias, para isso eles sempre registravam e guardavam as ocorrências, o último registro verificado desse banco de dados foi apresentado pelo trabalho da GIACOMITTI (2003) sob o tema “A estatística como fonte explicativa do trabalho de defesa social - Guarda Municipal de Curitiba”, este trabalho foi realizado sobre a base de ocorrências e apresentado análises com estatísticas descritivas a fim de contribuir nas estratégias de comunicação e proteção da Guarda Municipal de Curitiba.

## 2.2 LEI DE ACESSO À INFORMAÇÃO

A lei de acesso à informação garante o direito dos cidadãos de conseguir informações do governo disponibilizado de forma aberta a toda comunidade, aproximando mais o cidadão nos processos de tomada de decisão do governo e ajudando nas trocas de informações segundo o Senado Federal (2013).

A LAI estabelece que toda informação produzida ou custodiada por órgãos e entidades públicas é passível de ser ofertada ao cidadão, a não ser que esteja sujeita a restrições de acesso legalmente estabelecidas, quais sejam: (i) informações classificadas nos graus de sigilo reservado, secreto ou ultrassecreto, nos termos da própria LAI; (ii) informações pessoais, afetas à intimidade e à vida privada das pessoas naturais; ou (iii) informações

protegidas por outras legislações vigentes no País, como é o caso do sigilo fiscal e do sigilo bancário (Senado Federal, 2013)

No relatório de implementação da lei de acesso à informação do Senado Federal (2013) encontra-se que toda informação produzida por órgãos públicos é passível de ser ofertada para os cidadãos. A lei garante à sociedade acesso a informações em duas formas, a transparência passiva a qual é a disponibilização de conteúdo específicos solicitadas por pessoas físicas ou jurídicas e a transparência ativa a qual é fornecido de forma ampla dados de interesse geral da comunidade e disponibilizados principalmente em sítios eletrônicos.

## 2.3 GESTÃO DA INFORMAÇÃO

Rezende (2002, p. 78) afirma que estamos na era da sociedade do conhecimento, na qual as maiores fontes de riquezas vêm através da distribuição e manipulação da informação. Define-se que a inteligência empresarial é o processo de transformar dados em sabedoria conforme argumenta Rezende, e essa habilidade vem desde a coleta, organização, análise e implementação de mudanças a partir de informações e ele complementa que “Com a transição da Era da Informação para a Era do Conhecimento. Compreende-se que a informação, por si só, não gera novos conhecimentos” o que vai gerar o conhecimento são as interpretações e o compartilhamento.

Marchiori (2002, p.74) define a gestão da informação como um conjunto de processos que atuam desde o planejamento, organização, direção e distribuição de informação. O Gestor da Informação tem habilidades que ajudam a atender demandas informacionais, identificando as necessidades dos clientes, coletando e avaliando as informações, armazenagem das informações, distribuições dos conteúdos e sua utilização final.

Nóbrega, Bastos e Araújo (2014, p.43) por sua vez denominam a Gestão da informação (GI) um conjunto de etapas para a produção de informação, desde a organização, armazenamento, recuperação, acesso e uso da informação, auxiliando as organizações nos processos de tomada de decisão.

A forma em que as organizações usam a informação é estratégica e antes de se cumprir todas as etapas, a organização deve ter um entendimento além da tecnologia para ter sucesso na era da informação. De acordo com Davenport (1998,



p.18) é difícil definir informação, porque durante alguns anos as pessoas entendiam dado como informação. O Quadro 2 apresenta a diferença entre dado informação e conhecimento na visão de Davenport (1998, p.18)

QUADRO 2 - DIFERENÇA ENTRE DADOS, INFORMAÇÃO E CONHECIMENTO

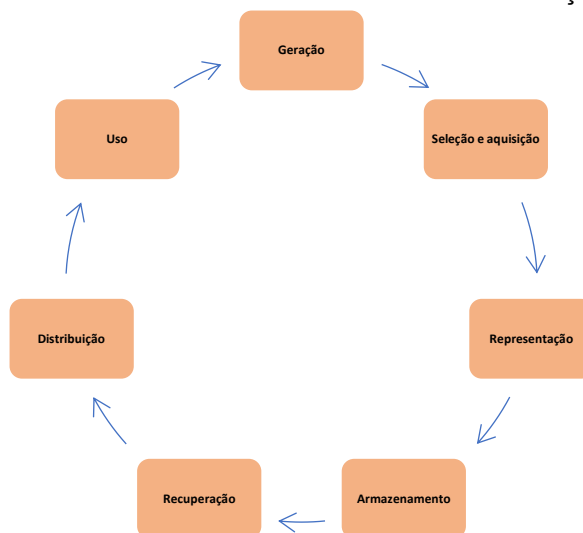
Dados	Informação	Conhecimento
<p>Simples observações sobre o estado do mundo</p> <p>Facilmente estruturado</p> <p>Facilmente obtido por máquinas</p> <p>Frequentemente quantificado</p> <p>Facilmente transferível</p>	<p>Dados dotados de relevância e propósito</p> <p>Requer unidade de análise</p> <p>Exige consenso em relação ao significado</p> <p>Exige necessariamente a mediação humana</p>	<p>Informação valiosa da mente humana</p> <p>Inclui reflexão, síntese, contexto</p> <p>De difícil estruturação</p> <p>De difícil captura em máquinas</p> <p>Frequentemente tácito</p> <p>De difícil transferência</p>

FONTE: Adaptado DAVENPORT (1998, p. 18).

Segundo-o Quadro 2, dados são facilmente obtidos por máquinas e é simples de serem observados, quando estes dados dotados de relevância e propósito em conjunto gera a informação, já o conhecimento é quando a informação é valiosa e difícil de estruturação ela é difícil de ser transferida comparado com os dados.

A gestão da Informação segue um ciclo de vida informacional atendendo todo um processo. Este ciclo informacional apresentado na Figura 1 é descrito por Tarapanoff (2006, p. 23) o qual destaca a busca das necessidades a fim de se obter informações a qual gera o tipo de informação necessária aumentando o conhecimento individual e coletivo das pessoas dentro da organização.

FIGURA 1 - MODELO DE FLUXO DE INFORMAÇÃO



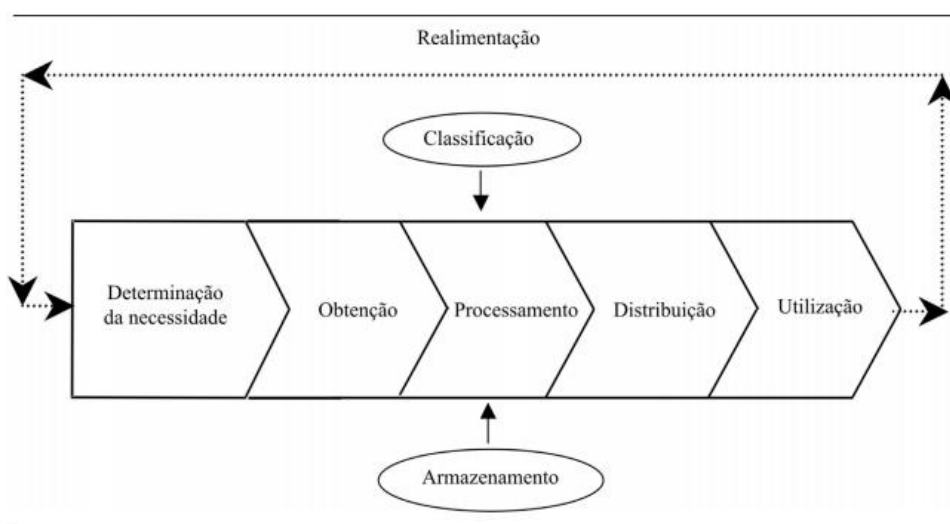
FONTE: TARAPANOFF (2006, p. 23)

Tarapanoff (2006, p. 23) apresenta através da Figura 1 o modelo de ciclo de informação, este ciclo está ligado desde a geração da informação, seleção, representação, armazenamento, recuperação, distribuição, uso e retorna para a geração.

Davenport (1998, p.173) considera que a Gestão da Informação dentro das organizações distribui e usa as informações. O conhecimento gerado é um dos pontos mais importantes de um processo informacional no qual é identificado as fontes de informações e os problemas dos usuários, e a partir deles indica os melhores caminhos para a tomada de decisão.

Moraes e Filho (2006, p.126) apresentam um processo de gerenciamento da informação que complementa o ciclo de vida da informação conforme apresentado na Figura 2. O processo de gerenciamento é um conjunto de atividades em que a captura as informações, distribui, utiliza e gera o conhecimento.

FIGURA 2 - O PROCESSO DE GERENCIAMENTO DA INFORMAÇÃO



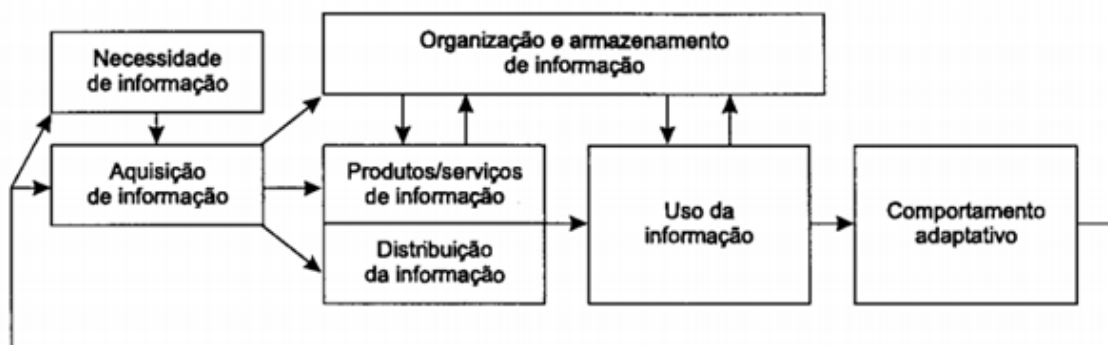
FONTE: MORAES; FILHO (2006, p.126)

O processo de gerenciamento de informações segundo Moraes e Filho (2006, p.125) é realizado em cinco etapas de acordo com a Figura 2. A primeira é a determinação das necessidades de informação dos clientes, identificando as fontes de informação necessárias para atender às principais demandas. A segunda é a obtenção, nesta etapa é feita a coleta de dados que serão seguidos nas outras atividades. A terceira parte do fluxo é o processamento, esta etapa é importante porque define a melhor forma de se acessar as informações e a forma que devem ser armazenadas. A quarta é a distribuição e apresentação da informação, quando são definidas as metodologias mais adequadas para apresentar e disponibilizar. A última parte do fluxo é a utilização, quando o cliente consome os resultados no processo estratégico e nas tomadas de decisão.

Choo (2003, p. 79) apresenta três observações sobre a coleta e análise de dados. Na primeira etapa, necessidades e uso de informação podem variar conforme a profissão e os usuários atendendo requisitos específicos. A segunda parte do fluxo traz que os usuários obtêm fontes de informações em ambientes formais e informais. E a terceira que as pessoas podem influenciar na seleção e uso das fontes. O autor destaca ainda a pergunta “Como as organizações usam a informação? ”, explicando que é uma tarefa difícil e que necessita de uma clara compreensão dos processos organizacionais. Pontua também que muitas empresas não são capazes de perceber a importância das suas fontes e tecnologia de informação CHOO (2003, p. 27).

Choo (2003, p. 126) aborda um conjunto com seis processos de informação, a qual se parece muito com o Processo de gerenciamento da informação fornecido por Moraes e Filho (2006, p. 126). O processo se chama ciclo de gestão da informação apresentado na Figura 3.

FIGURA 3 - CICLO DE GESTÃO DA INFORMAÇÃO



FONTE:CHOO (2003, p. 404)

O ciclo da Gestão da Informação, segundo Choo (2003, p.126), identifica as necessidades de informação seguindo para a aquisição da informação, organização e armazenamento da informação, desenvolvimento de produtos e serviços de informação, distribuição da informação, e uso da informação.

A Gestão da Informação retratada de acordo com Moreira (2014, p.17) é composta por três níveis hierárquicos distintos a quais os níveis são facilitadores na gestão da informação, obtendo resultados melhores em cada nível da organização apresentado na Figura 4.

FIGURA 4 - CICLO DE GESTÃO DA INFORMAÇÃO



FONTE: MOREIRA (2014, p.18)

O ciclo da gestão da informação de acordo com Moreira (2014, p.17) é composto por três níveis, o primeiro nível é o estratégico a qual é exclusivo da alta gerência, neste nível é feito as decisões que determinam os objetivos da organização. O segundo nível é o Tático chamado de administrativo, e ele está no nível do meio abaixo do estratégico e as decisões é tomado por gestores. O último nível, o operacional, é encarregado pelas tomadas de decisões de níveis mais baixos, a qual geralmente os responsáveis são os supervisores.

Cada vez mais as empresas têm que estar mais rápidas atendendo as necessidades e ao mesmo tempo dando qualidade de informação tanto para o ambiente interno como externo. A partir destes fatores, observou a necessidade das organizações trabalharem com suas informações melhorando seu ambiente de negócio.

A gestão da informação garante a gerência da informação de forma mais eficaz atendendo aos objetivos da empresa. Assis (2008, p.141) apresenta dez pontos básicos para a implementação da gestão da informação dentro das organizações que são:

“A gestão da informação deve estar alinhada com a missão e os objetivos estratégicos; Desenvolver um plano de gestão da informação voltado, preferencialmente, para a perspectiva do negócio; Preocupar-se sempre com a máxima: a informação para as pessoas certas, no local correto, no tempo

certo, no formato adequado e, se possível, com custo zero; Ter sempre a visão de que a informação deve ser utilizada no seu potencial máximo; Priorizar a qualidade, a disponibilidade, o uso e o valor da informação; O gestor da informação deve estar ligado diretamente à alta administração; Mapear regularmente as necessidades de informação; Considerar a qualidade das fontes de informação e sua disponibilidade; Permanentemente, analisar o custo x benefício das fontes de informação adquiridas; Contextualizar e compartilhar a informação de interesse” (ASSIS, 2008, p. 141).

Através destes dez pontos destacados por Assis (2008), o autor pretende atender pontos como alinhamento aos objetivos estratégicos, análise de fontes de informação, compartilhamento de informações e o custo-benefício.

### 2.3.1 Recuperação da informação

Pontes, Carvalho e Azevedo (2013) a recuperação da informação (RI) é um campo que abrange diversos domínios e possui ferramentas de organização e recuperação como tesauros, taxonomia e ontologias (dentre outras) estabelecendo diversos estudos e visões.

A recuperação da informação localiza documentos e itens previamente armazenados, através de extração e síntese de conceitos, demanda dos usuários e de termos existentes nas bases de dados Figueiredo (2006, p. 19).

Bonassa (2009) apresenta que a recuperação da informação requer um processo de organização dos dados de forma criteriosa e que necessita de tratamento técnico na informação. O desenvolvimento de estratégias para recuperação da informação auxilia na localização e a partir das informações recuperadas produz-se o conhecimento. A recuperação investiga algo perdido através de um arranjo ordenado de conhecimentos.

A recuperação de informação é o nome para o processo ou método pelo qual um usuário potencial de informação é capaz de converter a sua necessidade de informação em uma lista real de citações de documentos armazenados contendo Usabilidade na recuperação da informação em catálogos bibliográficos informações úteis para ele. É o processo de descoberta a respeito da informação armazenada. No mais, é o nome para a produção de uma bibliografia em demanda. [...] A recuperação de informação é crucial para documentação e organização do conhecimento. (BONASSA, 2009, p. )

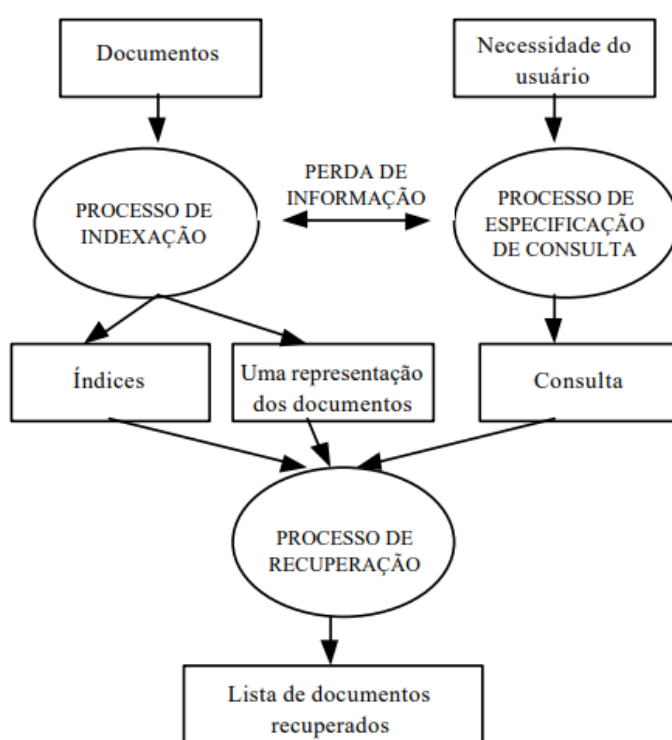
A recuperação da informação é um processo em que transforma a necessidade da informação em algo plausível e documentado, ele descobre informações armazenadas e é fundamental para a organização do conhecimento (BONASSA, 2009).



Assis (2008, p. 143) aponta que o grande crescimento e volume de publicações pelo mundo levou à necessidade de suprir as necessidades dos usuários.

De acordo com Cardoso (2004, p. 1) não se fala apenas em crescimento da informação, mais sim em explosão informacional, com este grande volume cada vez fica mais complexo recuperar informações e é se exigido mais em processos. Através deste contexto o autor apresenta níveis hierárquicos organizacionais conforme a Figura 5.

FIGURA 5 - NÍVEIS HIERÁRQUICOS ORGANIZACIONAL



FONTE: CARDOSO (2004, p.1)

O fluxo de níveis hierárquicos organizacional apresentado na Figura 5 se inicia com a inclusão dos documentos e com as necessidades dos usuários gera as consultas recuperando listas relevantes de conteúdo e o processo de especificação das consultas geralmente é difícil segundo a Cardoso (2004, p. 1).

### 2.3.2 Tomada de decisão

Segundo Bertoni, Brito, Silva e Leme (s.d.), a tomada de decisão é presente diariamente dentro das organizações tanto dos administradores quanto dos funcionários em geral, ele destaca que não existe uma decisão perfeita, mas sim a busca pela melhor alternativa a qual ajudará a organização a se obter o sucesso. A tomada de decisão e a resolução de problemas dentro da organização podem ser abordadas de muitas maneiras diferentes, mas geralmente seguem dois modelos o racional e comportamental conforme o Quadro 3.

QUADRO 3 - MODELO RELACIONAL E COMPORTAMENTAL DE TOMADA DE DECISÃO

Modelo Racional	Modelo Comportamental
1. O tomador de decisões tem informações perfeitas (relevantes e acuradas).	1. O tomador de decisões tem informações imperfeitas (incompletas e possivelmente imprecisas)
2. O tomador de decisões tem uma lista exhaustiva de alternativas dentre as quais pode escolher.	2. O tomador de decisões não tem um conjunto completo de alternativas ou não entende plenamente aquelas que têm à disposição.
3. O tomador de decisões é racional	3. O tomador de decisões tem uma racionalidade definida e se restringe a valores, experiência, hábitos etc.
4. O tomador de decisões sempre tem em mente os melhores interesses da organização.	4. O tomador de decisões escolherá a primeira alternativa minimamente aceitável.

FONTE: CARAVANTES; PANNO; KLOECKER (2005, p.455)

O Quadro 3 apresenta as diferenças entre o modelo relacional e o modelo comportamental, o modelo racional é mais favorável para tomar as melhores decisões pelo fato que se tem informações mais completas e perfeitas, ele tem lista de alternativas para escolher, ele é racional e ele sempre tem as decisões pensando no melhor para a organização. Já o modelo comportamental tem poucas informações, não tem muitas alternativas de escolhas para decisões, tem uma racionalidade mais restritiva e aceita a primeira alternativa mais aceitável.

O modelo racional supõe que quem toma as decisões tem informações perfeitas, além de avaliar de modo sistemático e lógico para cada alternativa decidindo

qual a melhor para a organização. Já o modelo comportamental está relacionado a tomada de decisão que leva em conta sua percepção, experiência, informações e alternativas limitadas de acordo com Bertoni, Brito, Silva e Leme (s.d.).

De acordo com Chiavenato (2003, p.347) a teoria das decisões surgiu com Herbert Simon, que a utilizou como base para explicar o comportamento humano nas organizações. Esta teoria traz que o comportamento gera o processo de tomada de decisões na organização e esta teoria comportamental diz que o administrador não é o único que toma decisões, mais sim a organização toda. E ele ressalta que “a organização é um complexo sistema de decisões”.

Existem várias teorias sobre as diferentes maneiras do comportamento humano nas organizações e Chiavenato (2003, p.348) apresenta três delas: a Teoria clássica da administração, a Teoria das relações humanas e Teoria comportamental.

A teoria clássica da administração traz que as pessoas dentro da organização são instrumentos passivos, a qual tem uma posição simplista e mecanicista além de que pode haver um aumento nas produtividades dentro da organização com incentivos fiscais.

A teoria das relações humanas defende que as pessoas dentro da organização são possuidores de necessidades, atitudes, valores e objetivos pessoais a qual precisam obter participação dentro da organização através de estímulos e serem compreendidos.

Já a teoria Comportamental está relacionada à decisão das pessoas dentro da organização, estabelecendo que as mesmas raciocinam e decidem se participam ou não na tomada de organização dentro da organização.

Portanto, Chiavenato (2003, p.348) conclui que estas três teorias classificam pessoas dentro da organização e determina cada tipo de pessoa. O autor apresenta seis elementos nos quais o processo de tomada de decisão é executado. O primeiro é a decisão dos tomadores de decisão, os objetivos a qual o tomador deve alcançar, as preferências e os critérios executados por cada tomador, a estratégia abordada a qual ajuda a atingir os objetivos propostos, a situação a qual identifica aspectos do ambiente em que se envolve e por fim o resultado a qual é a consequência de uma estratégia.

A organização é um sistema de decisões em que cada pessoa participa consciente e racionalmente, escolhendo e decidindo entre alternativas mais ou menos racionais que são apresentadas de acordo com sua personalidade, motivações e atitudes. Os processos de percepção das situações e o raciocínio são básicos para a explicação do comportamento humano nas organizações: o que uma pessoa aprecia e deseja influencia o que se vê e interpreta, assim como o que vê e interpreta influencia o que aprecia e deseja. Em outros termos, a pessoa decide em função de sua percepção das situações. Em resumo, as pessoas são processadores de informação, criadoras de opinião e tomadores de decisão. (Chiavenato, 2003, p.348)

O autor ainda ressalta que a organização tem participação de todos os colaboradores no processo de tomada de decisão, se baseiam na interpretação e além das de decisões ele cria opiniões.

Maximiano (2000, p.139) apresenta que as decisões são escolhas a qual as pessoas fazem para solucionar problemas e aproveitar oportunidades. O processo de tomada de decisão atende vários tipos de decisões e o autor apresenta as principais formas de classificar as decisões conforme o Quadro 4:

QUADRO 4 - PRINCIPAIS TIPOS DE DECISÕES

Tipo de decisões	Formas para classificar
Programadas	Aplicam-se a problemas repetitivos
Não programadas	aplicam-se a problemas não familiares
Estratégicas	Escolhem objetivos para a organização
Administrativas	Colocam decisões estratégicas em prática
Operacionais	Definem meios e recursos
Individuais	São tomadas unilateralmente
Coletivas	são tomadas em grupos
Satisfatórias	qualquer alternativa serve
Maximizadas	procuram o melhor resultado possível
Otimizadas	equilibram vantagens e desvantagens de diversas alternativas

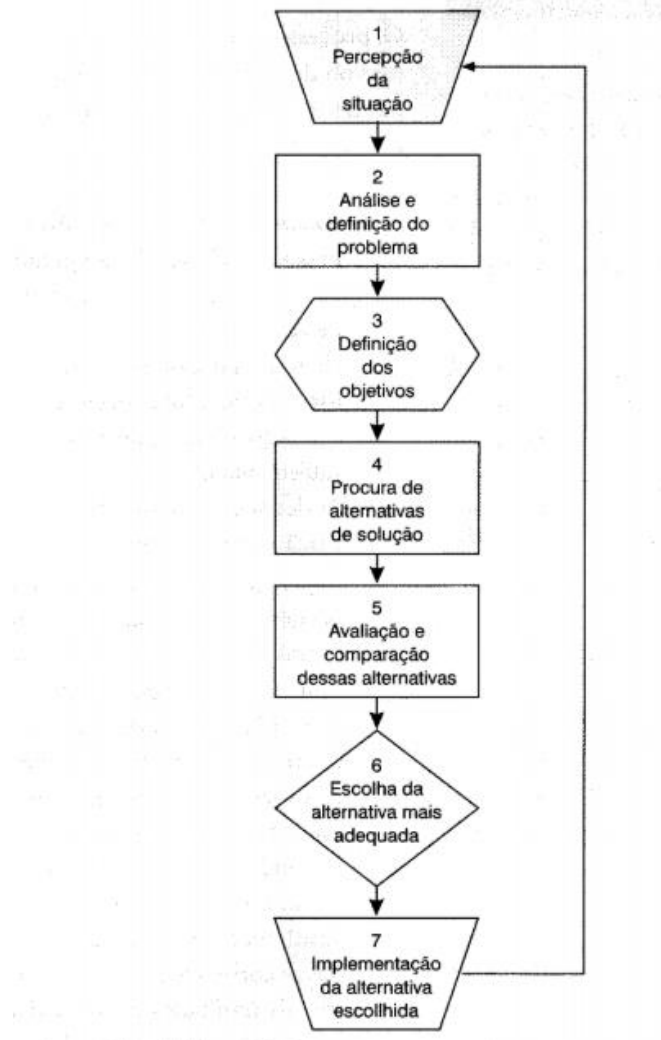
FONTE: MAXIMIANO (2000, p.139)

Estes tipos de decisões apresentadas no Quadro 4 atendem diversos fatores para a tomada de decisão, auxiliando o tomador de decisão com classificações perante a cada tipo de situação.

Para o processo decisório existem sete etapas e para a sua elaboração são necessárias características pessoais do tomador de decisões, da situação em que

está envolvido e da maneira como percebe a situação cada etapa é representada conforme a sequência na Figura 6.

FIGURA 6 - PROCESSO DECISÓRIO SOB A ÓTICA DA ORGANIZAÇÃO

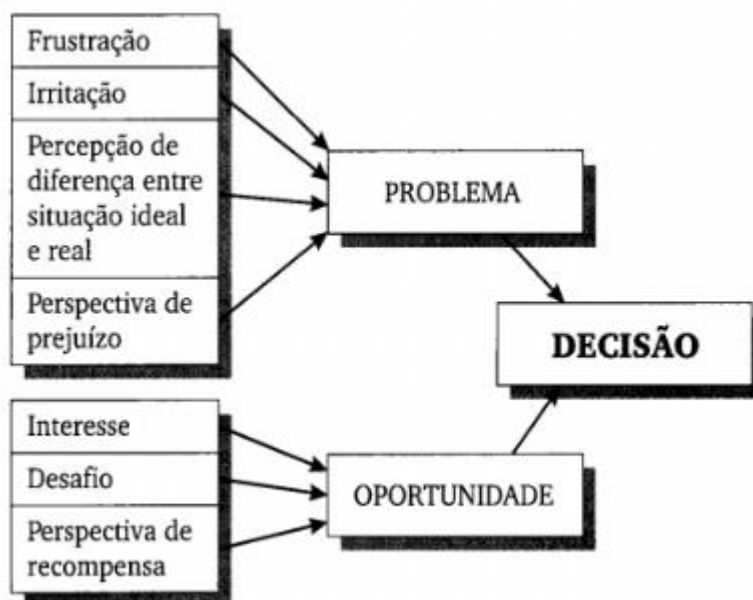


FONTE: CHIAVENATO (2003, p.350)

Com o processo decisório dentro da organização pode haver várias decorrências de acordo com Chiavenato (2003, p. 349) como uma racionalidade limitada, imperfeição na decisão, relatividade nas decisões, hierarquização das decisões, racionalidade administrativa e a Influência organizacional. Durante o processo decisório o tomador de decisões evita incertezas e segue regras padronizadas pela organização.

Maximiano (2000, p. 139) apresenta que os problemas podem ser definidores de diversas maneiras conforme apresentado na Figura 7, tanto por frustração tanto por oportunidades.

FIGURA 7 - DECISÕES SÃO ESCOLHAS QUE PROCURAM RESOLVER PROBLEMAS E APROVEITAR OPORTUNIDADES



FONTE: MAXIMIANO (2000, p. 140)

De acordo com Maximiano (2000, p. 139), a Figura 7 apresenta problemas e oportunidades que influenciam na tomada de decisão, a decisão é uma escolha do decisor com o intuito de resolver e aproveitar as oportunidades.

Segundo Bertoncini, Brito, Silva e Leme. (s.d.) existem técnicas para auxiliar no processo de tomada de decisão a qual ajuda os administradores a melhorar a decisão dos problemas, a partir da utilização de análise de vantagens e desvantagens, árvore de decisões (representação gráfica das alternativas, análise do campo de força possibilitando analisar forças a favor e forças contrária auxiliando as decisões e ponderação de critérios afim de possibilitar avaliação de fatores que atendem as necessidades e preferências dos tomadores de decisão.

## 2.4 KDD (KNOWLEDGE DISCOVERY IN DATABASE)

O *Knowledge discovery in database* (KDD) ou traduzido como “descoberta de conhecimento em bases de dados” é um processo que tem por finalidade descobrir conhecimento em base de dados. Apresenta as etapas de obtenção de dados, pré-processamento, mineração de dados e pós processamento de acordo com Silva, Peres e Boscarioli (2016, p. 11). Um dos principais objetivos da descoberta de conhecimento é encontrar padrões intrínsecos aos dados, facilitando a assimilação para o conhecimento. A Figura 8 contém o processo de geração de conhecimento.

FIGURA 8 - PROCESSO KDD



FONTE: Adaptado SILVA; PERES; BOSCAROLI (2016, p.11)

De acordo com Silva, Peres e Boscarioli (2016, p. 11) o KDD se inicia com a obtenção dos dados, nesta fase são organizados os dados relacionados com a área de interesse e são necessários o conhecimento e domínio da área para seguir com o processo.

A segunda fase do KDD o pré-processamento é uma das fases mais importantes do processo, segundo Silva, Peres e Boscarioli (2016, p. 11) no pré-processamento é realizada a organização do repositório, com a limpeza dos dados,

tratamento da base de dados com eliminação de instâncias repetidas, inconsistências, valores discrepantes, valores ausentes e ruidosos.

A terceira fase do KDD é a Mineração de Dados, este processo é realizado após o pré-processamento, etapa em que se aplicam os algoritmos com o objetivo de extrair conhecimento, envolvendo a resolução de tarefas tais como predição, agrupamento, associação ou detecção de anomalias de acordo com o Silva, Peres e Boscarioli (2016, p. 11).

Por fim a quarta etapa é o pós processamento e Silva, Peres e Boscarioli (2016, p. 11) define que esta etapa analisa os resultados obtidos através da mineração de dados e estatística, gerando o conhecimento.

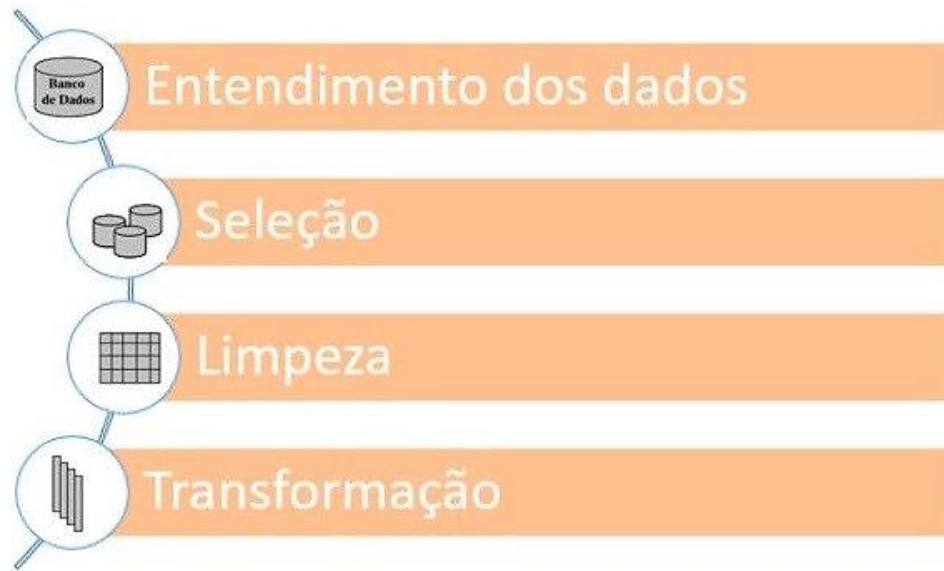
Ainda segundo os autores, estes processos não precisam necessariamente seguir uma sequência, ele pode voltar para uma etapa várias vezes mesmo estando em outra.

#### 2.4.1 Pré-processamento

O pré-processamento como dito anteriormente no KDD é uma das fases principais do KDD, e é composto pela fase de entendimento dos dados, seguindo para a seleção, limpeza e transformação dos mesmos. Neves (2003, p. 31) destaca que não existe uma sequência obrigatória para este processo, podendo ser realizado da matéria que melhor atribuir para a base selecionada para o estudo, na Figura 9 contém os processos para a realização do pré-processamento.



FIGURA 9 - SUBFASES DO PRÉ-PROCESSAMENTO

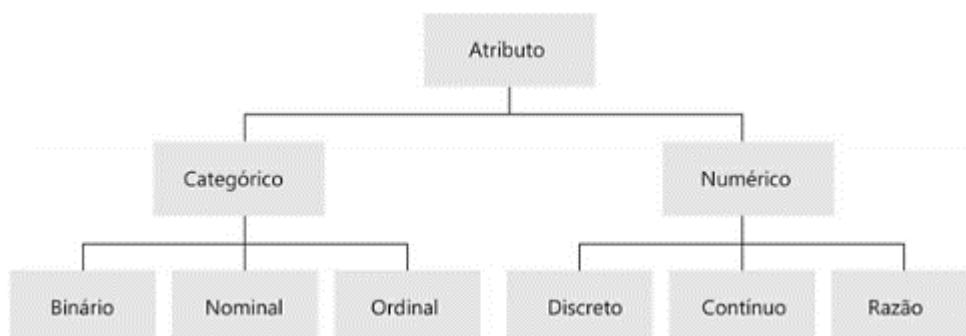


FONTE: Adaptado NEVES (2003, p. 31)

Através da realização das subfases do pré-processamento apresentado na Figura 9 é possível garantir uma melhor qualidade para os dados, através de métodos com padronização, tratamento de valores e eliminação de dados duplicados e errôneos, faz com que o resultado do KDD seja mais eficiente e exato.

O primeiro passo do pré-processamento é a seleção dos dados, Neves (2003, p. 32) explica que nesta etapa são escolhidas as instâncias e atributos da tabela relacionados com as necessidades e objetivos do pesquisador, e para isso é necessário se ter um entendimento dos dados. Na Figura 10 apresenta a relação dos tipos de atributos analisados na base de dados.

FIGURA 10 - TIPOS DE ATRIBUTOS



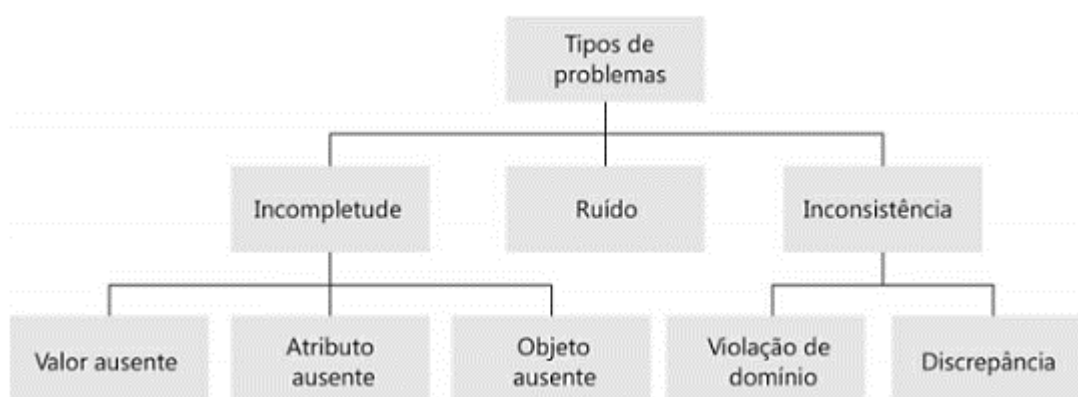
FONTE: CASTRO; FERRARI (2016, p. 71)

O entendimento dos dados é a primeira tarefa para iniciar a limpeza e transformação dos mesmos, é realizado um levantamento dos atributos e instâncias identificando o tipo de cada atributo de acordo com a Figura 10. A importância do atributo impacta na escolha dos algoritmos e em relação às suas principais necessidades.

Segundo Neves (2003, p. 32) após o entendimento da base é realizada a seleção dela, nesta etapa é realizado junto ao usuário de modo que atende às principais necessidades para selecionar as tabelas, atributos, e instâncias gerando um novo conjunto de dados para ser submetido às demais fases dos sub processos do pré-processamento.

A terceira etapa segundo Castro e Ferrari (2016, p. 67) é a limpeza dos dados selecionados, nesta etapa são analisados os tipos de valores em cada atributo. É realizado também a verificação da qualidade dos dados verificando se há dados faltantes, escritos de forma errada, dados duplicados e com falta de padronização. Segundo os autores eles classificam os principais tipos de problemas identificados em uma base de dados conforme a Figura 11.

FIGURA 11 - PRINCIPAIS PROBLEMAS COM DADOS



FONTE: CASTRO; FERRARI (2016, p. 67)

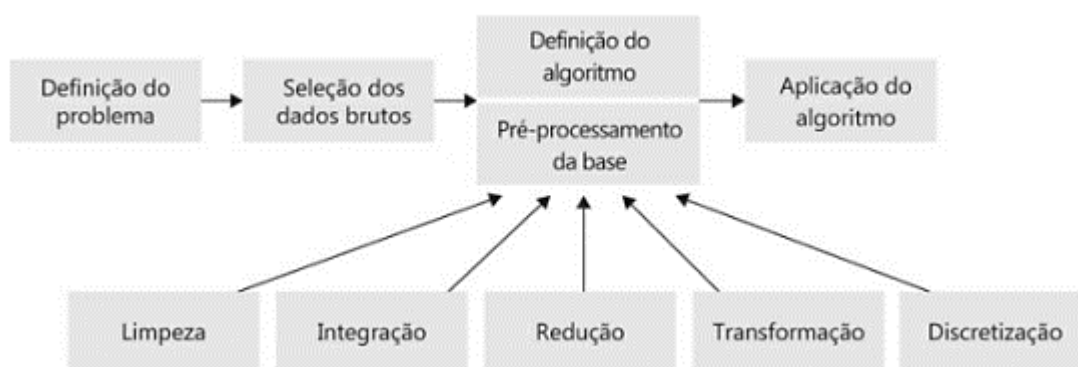
Castro e Ferrari (2016, p. 67) distinguem os tipos de problemas conforme a Figura 11, para eles existem três tipos de problemas com os dados. A incompletude é um deles, neste problema podem apresentar valores ausentes em atributos e objetos. No tipo de problema inconsistência é o conflito dos dados que aparecem em locais

diferentes e dados discrepantes. E por fim existem os ruídos que podem causar inconsistências na base.

A última etapa do pré-processamento é a transformação, Neves (2003, p. 60) explica que nesta etapa é realizado a transformação de alguns dados para aplicação dos algoritmos selecionados, esta etapa pode ser em relação a normalização dos dados a qual consiste na conversão de atributos, há também a desratização a qual é a transformação de valores contínuos para valores discretos e vale para formação de categorias de valores que já são discretos, além de outras transformações como a conversão de valores simbólicos para numéricos, agrupamentos e composição de atributos.

A fase de pré-processamento do KDD é a mais trabalhosa como já dito antes por Castro e Ferrari (2016), e para se conseguir os melhores resultados dos algoritmos com qualidade, depende de um bom pré-processamento da base de dados, atendendo os requisitos de cada algoritmo e necessidades dos usuários. Através da Figura 12 contém as etapas já discutidas anteriormente da preparação da base de dados do Castro e Ferrari (2016, p. 78).

FIGURA 12 - ETAPAS DO PROCESSO DE PREPARAÇÃO DA BASE DE DADOS



FONTE: CASTRO; FERRARI (2016, p. 78)

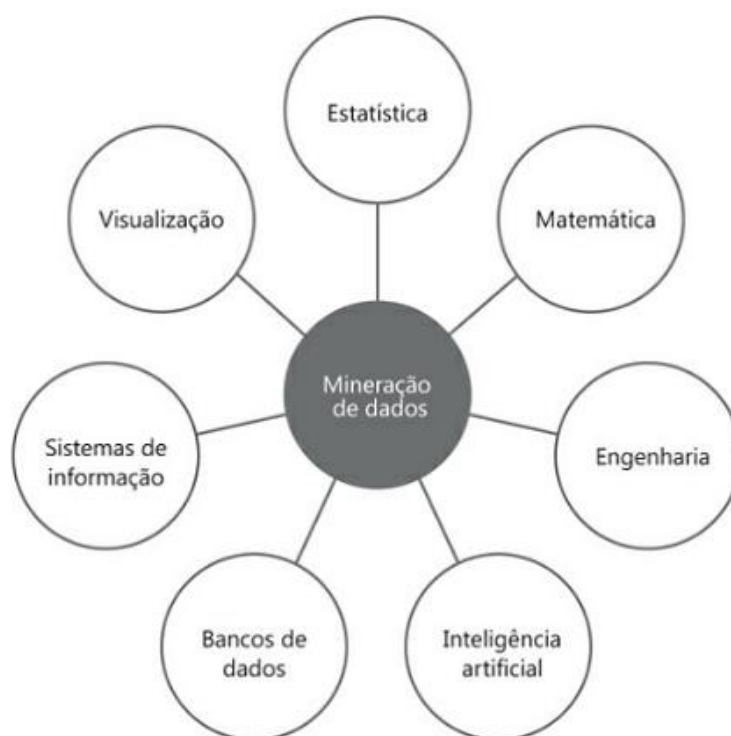
Através das etapas dos processos de preparação da base de dados observado na Figura 12 no pré-processamento é realizada toda a limpeza de dados inconsistentes, ruidosos, ausentes e redundantes, além de ser realizado a integração entre os atributos e instancias, reduzindo massa de dados, transformando e adaptando a base para ser aplicado algoritmos e realizado discretizações.

### 2.4.2 Mineração de dados

A mineração de dados ou *Data Mining* em inglês é definida segundo Silva, Peres e Boscardoli (2016, p. 11) em termos de esforço para a descoberta de padrões em bases de dados ajudando nas tomadas de decisões. De forma geral trata-se de aplicações de algoritmos e técnicas estatísticas que recebe um conjunto de dados e traz como saída um padrão de comportamento que podem ser expressos como uma regra de associação, função de mapeamento ou uma modelagem de perfil. A mineração de dados trata de uma parte do KDD “descoberta de conhecimento” e na literatura especializada o autor explica que são encontrados diversos tipos de taxonomias que caracterizam as tarefas de mineração de dados.

Castro e Ferrari (2016, p. 39) apresenta que a a mineração de dados envolve vários conhecimentos em áreas distintas conforme apresentado na Figura 13.

FIGURA 13 - MULTIDISCIPLINARIDADE DA MINERAÇÃO DE DADOS



FONTE: CASTRO; FERRARI (2016, p. 39).

De acordo com a Figura 13 a mineração de dados compõe vários conhecimentos ela é multidisciplinar e divide áreas como banco de dados, estatística, aprendizagem de máquina, computação de alto desempenho, reconhecimento de

padrões, computação natural, visualização de dados, matemática, inteligência artificial e entre outros.

Conforme apresentado anteriormente, a mineração é composta por tipos de taxonomias e uma das apresentadas pelo autor Fayyad et al. (1996) compõe em dois níveis a tarefa de mineração de dados divididas entre preditivas e descritivas. Nas tarefas preditivas são utilizados valores descritivos com a função de prever valores futuros ou desconhecidos os autores inserem nesta taxonomia algoritmos de classificação e regressão. Já nas tarefas descritivas apresenta-se que esta tem como objetivo detectar padrões que melhor descrevem os dados para interpretação e no conjunto de dados nas tarefas descritivas é apresentado algoritmos de agrupamento, sumarização, modelagem de dependências e detecção de anomalias. No Quadro 5 está a representação destas técnicas mais comuns:

QUADRO 5 - TÉCNICAS DE MINERAÇÃO DE DADOS

<b>Mineração de Dados</b>	<b>Preditivas</b>	Classificação
		Regressão
	<b>Descritivas</b>	Agrupamento
		Associação
		Detecção de Anomalias
		Sumarização

FONTE: A autora (2019)

Dentro deste contexto de técnicas de mineração de dados, de acordo com o Quadro 5, nas próximas sessões serão apresentadas as técnicas de mineração e seus algoritmos correlacionados dentro do quadro de descritivas e preditivas.

A mineração segundo Silva, Peres e Boscardoli (2016, p. 40) é realizada através de duas técnicas, a técnica descritiva e as técnicas preditivas. As técnicas descritivas segundo Silva, Peres e Boscardoli (2016, p. 40) identificam padrões de comportamento comum e caracterizam as propriedades gerais dos dados. De forma geral ela mede, explora e descreve características intrínsecas dos dados, e com estas análises é

possível investigar distribuições de frequência, medidas de centro de variação, medidas de posição relativa e associação dos dados. Em relação a técnica descritiva existe as seguintes heurísticas apresentadas no Quadro 6.

QUADRO 6 - HEURÍSTICAS DA TÉCNICA DESCRITIVA

Heurísticas	Descrição
Agrupamento	As técnicas de agrupamento têm a finalidade de identificar grupos de instâncias a partir de seus atributos. Estes grupos são conhecidos como Clusters e eles maximizam as similaridades do dado dentro do grupo.
Associação	As técnicas de associação têm como finalidade encontrar relações ou dependências entre as instâncias de um atributo da análise do conjunto de dados.
Deteccção de Anomalias	As técnicas de deteção de anomalias ou também conhecida como deteção de outliers, ajuda na identificação de anomalias nos dados, removendo dados discrepantes que não enquadram no comportamento dos demais registros.
Sumarização	As técnicas de Sumarização é a generalização dos dados, ela extrai informações condensadas de uma massa de volume de dados e fornece um resultado com um conjunto menor, tendo uma visão mais geral dos dados.

FONTE: A autora (2019)

Conforme apresentado no Quadro 6 cada heurística é um grupo que atende determinada situação, e dentro de cada grupo há vários algoritmos. Por exemplo na heurística de agrupamento tem o algoritmo k-Means, na heurística de associação têm o algoritmo de a priori, o de deteção de anomalias têm as redes neurais e de sumarização tem o Luhn.

Outra técnica da mineração de dados são as preditivas a qual contempla técnicas de classificação e regressão. A técnica preditiva faz inferência a partir de dados objetivando predições e utiliza valores dos atributos descritivos para prever valores futuros ou desconhecidos de outros atributos segundo Silva, Peres e Boscarioli (2016, p. 40). No Quadro 7 contém a descrição de suas duas heurísticas.

QUADRO 7 - HEURÍSTICAS DA TÉCNICA PREDITIVA

Heurísticas	Descrição
Classificação	Esta técnica busca encontrar modelos de predição para variáveis meta do tipo contínuas e ela utiliza todos os atributos que compõem para predisse os resultados.
Regressão(estimação)	Já a técnica de regressão também prediz o resultado e ela tem como classe numérica.

FONTE: A autora (2019)

Conforme apresentado no Quadro 7 as heurísticas atendem um método de análise adaptado a base, e dentro de cada grupo há vários algoritmos. Na heurística de classificação temos como exemplo o algoritmo de árvore de decisão e Naïve Bayes e na de regressão o algoritmo de regressão linear.

#### 2.4.3 Pós-processamento

O pós-processamento, de acordo com Milani e Carvalho (2013), serve para eliminar a redundância, generalizar, identificar no conjunto aqueles com maior potencial de serem interessantes de forma que contribui para a solução do problema encontrado.

Já segundo Goldschmidt e Passos (2005, p. 55) o pós-processamento é a última fase do KDD, e nele são apresentados todos os resultados e analisados com o objetivo de definir novas formas de investigação de dados.

De forma geral, esta etapa envolve a visualização da análise segundo Boente, Oliveira e Rosa (s.d., p. 9) e a interpretação da mesma, em relação ao processo de mineração de dados o pós-processamento auxilia na simplificação do modelo de conhecimento, organização e apresentação dos Resultados.

### 3 METODOLOGIA

O presente capítulo tem como objetivo apresentar os elementos iniciais do desenvolvimento deste estudo, tais como: caracterização da pesquisa; ambiente de pesquisa; dicionário de dados e os métodos utilizados durante o processo do KDD sendo (pós-processamento, estatística e mineração de dados).

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

A definição da caracterização de pesquisa deve analisar a finalidade da pesquisa, abordagem, objetivos e os procedimentos. Seguindo este raciocínio Gil (2008, p. 26) aponta que cada pesquisa tem um objetivo específico, porém pode ser agrupado a várias pesquisas. Ele explica que a finalidade da pesquisa é classificada em uma pesquisa pura ou uma pesquisa aplicada, e nesse caso, o estudo é realizado através de uma pesquisa aplicada, a fim de responder questões específicas com o objetivo de soluções concretas e resultados, através das descobertas geradas durante o presente estudo enriquece o conhecimento deste trabalho.

O direcionamento do trabalho pode ser definido como Análise Descritiva, Análise Preditiva e Análise Prescritiva. A base de dados da Guarda Municipal contém dados de ocorrência registradas e para esta pesquisa foi utilizada a análise descritiva.

Para o desenvolvimento deste trabalho foi utilizado o objeto de pesquisa descritivo o qual vai ajudar a descrever as características dos atributos da base entre as variáveis e nela são utilizadas técnicas a fim de analisar os fenômenos sem interferência do pesquisador.

A última etapa da caracterização da pesquisa é a definição dos procedimentos dentre os quais existem oito tipos, de acordo com Silva e Menezes (2005, p. 21), o procedimento realizado para a elaboração deste estudo é a pesquisa experimental a qual consiste em um objeto de estudo em que é capaz de influenciar na sua produção e ela tem como base a tentativa e o erro, e com a utilização dela vai ser possível compreender a causa dos fenômenos produzidos sobre a base.

Seguindo esta caracterização de pesquisa foram definidas as classificações da pesquisa conforme a Quadro 8, destacando os de cor laranja aqueles utilizados para a realização do presente trabalho.



QUADRO 8 - CLASSIFICAÇÃO DAS PESQUISAS

Finalidade	Análise	Objetivos	Procedimento
Aplicada	Descritiva	Descritiva	Bibliográfica
Pura	Preditiva	Explicativa	Documental
	Prescritiva	Exploratória	Estudo de caso
			<b>Experimental</b>
			Levantamento
			Participante
			Expost-Facto
			Pesquisa-Ação

FONTE: A autora (2020)

### 3.2 AMBIENTE DE PESQUISA

Este trabalho foi realizado na base de dados da Guarda Municipal de Curitiba no seguimento de Segurança Pública Municipal.

A Guarda Municipal presta serviços para a capital do Paraná, município de Curitiba, e atende a diversos tipos de ocorrências tais como: música alta, alarmes, roubos, invasão alagamentos, danos e entre outros. Por ser um órgão público a base de dados está disponível no site da prefeitura de Curitiba na sessão de dados abertos, disponível para todos os cidadãos conforme a lei de acesso à informação, facilitando o acesso para o desenvolvimento desta pesquisa.

### 3.3 DICIONÁRIO DE DADOS

A base de dados da Guarda Municipal de Curitiba contém 280.905 registros de ocorrências atendidas do ano de 2009 até o mês 5 de 2020 e 32 atributos. A base de dados encontra-se no formato de CSV disponível no site<sup>1</sup> da prefeitura de Curitiba na seção de dados abertos e os seus principais atributos estão descritos segundo o Quadro 9 separados pelo tipo de atributo.

<sup>1</sup> Fonte: <http://www.curitiba.pr.gov.br/dadosabertos/>

QUADRO 9 - QUADRO DE ATRIBUTOS DA BASE DE DADOS DA GUARDA MUNICIPAL DE CURITIBA

(continua)

Atributo	Valor do atributo	Tipo de atributo	Descrição
ATENDIMENTO_ANO	Variável Ex: 2009	Inteiro	Ano da ocorrência Realizada
ATENDIMENTO_BAIRRO_NOME	Variável: EX: BACACHERI	Nominal	Nome do bairro em que foi realizado o atendimento
FLAG_EQUIPAMENTO_URBANO	Variável: SIM NÃO	Nominal	Flag para identificar se é um equipamento urbano ou não
FLAG_FLAGRANTE	Variável: SIM NÃO	Nominal	Flag para identificar se é um equipamento urbano ou não
NATUREZA1_DEFESA_CIVIL	Variável: 0 1	Inteiro	Flag para identificar se é uma natureza de defesa civil ou não (0 - Não, 1 - Sim)
NATUREZA1_DESCRICAO	Variável: ex: Perturbação do sossego	Nominal	Descrição da primeira natureza cadastrada na ocorrência
OCORRENCIA_ANO	Variável Ex: 2009	Inteiro	Ano de cadastro da ocorrência
OCORRENCIA_CODIGO	Variável Ex: 11516	Inteiro	Código da ocorrência
OCORRENCIA_DATA	Variável Ex: 2009-01-23 11:17:00.000 11516	Date	Data da ocorrência
OCORRENCIA_DIA_SEMANA	Variável: EX: SEGUNDA	Nominal	Dia da semana em que a ocorrência foi cadastrada
OCORRENCIA_HORA	Variável: EX: 19:45:00	Time	Hora em que a ocorrência foi cadastrada
OCORRENCIA_MES	Variável: EX: 1	Inteiro	Mês em que a ocorrência foi cadastrada
ORIGEM_CHAMADO_DESCRICAO	Variável: EX:CIOSP (190)	Alfanumérico	Local em que se originou a chamada

REGIONAL_FATO_NOME	Variável: EX: BOA VISTA	Nominal	Regional do fato (local em que ocorreu a ocorrência)
SECRETARIA_SIGLA	Variável: EX:FAS	Nominal	Sigla da secretaria solicitante

FONTE: SiGesGuarda (2020)

Através do Quadro 9 com a descrição de cada atributo da base de dados é possível logo analisar um exemplo de cada atributo, o tipo de dados e a descrição de cada atributo disponibilizados no site com a base de dados disponível.

### 3.4 PRÉ PROCESSAMENTO

Através do pré-processamento, neste trabalho foi realizada a limpeza dos dados e o tratamento das bases.

Inicialmente antes de começar a limpeza dos dados foi realizado um levantamento com o total de ocorrências na base de dados conforme descrito no capítulo 3.3 Dicionário de dados, a base de dados contém 35 atributos e 280.905 registros de ocorrências do ano de 2009 até o mês 5 de 2020.

#### 3.4.1 Limpeza dos dados

A limpeza da base de dados segundo Silva, Peres e Boscardoli (2016, p. 42 a 51) é considerada a partir de quatro tipos de validações conforme o Quadro 10.

QUADRO 10 - RELAÇÃO DE VALIDADORES DA BASE DE DADOS

<b>Validação</b>	<b>Descrição</b>
Valores ruidosos	Os valores ruidosos ou conhecidos como outliers, são valores modificados em relação ao original a qual é diferente entre os outros esta é a primeira validação a ser analisada neste trabalho.
Valores inconsistentes	A segunda validação realizada é relacionada a valores Inconsistentes, valores com discrepâncias para um mesmo tipo de atributo.
Valores ausentes	A terceira validação é relacionada a dados ausentes, e neste caso ele está associado a dois tipos de fatores, o primeiro em relação a falta de valores para um atributo e o segundo quando um atributo depende destes valores ausentes.
Valores redundantes	A quarta e última validação está relacionado a valores com redundância, neste caso pode ocorrer por três fatores o primeiro é o uso de nomenclatura diferentes, o segundo relacionado armazenamento de atributos derivados e o último a inserção de valores repetidos.

FONTE: A Autora (2020)

Através destas quatro análises apresentadas no Quadro 10, todas foram utilizadas a tabela dinâmica do Excel como ferramenta de pesquisa e eliminação de dados, no capítulo (4.1.1 Limpeza dos dados) contém de forma detalhada cada análise realizada neste trabalho.

### 3.4.2 Tratamento da base de dados

O tratamento da base de dados é a última fase realizada no pré-processamento deste trabalho, através do tratamento à base de dados é ajustada para que possa ser realizado estatísticas descritivas e aplicação dos algoritmos de mineração de dados. Foram utilizados o Excel e o R Studio como ferramentas de ajustes da base conforme descrito no capítulo (4.1.3 Tratamento da Bases de Dados).

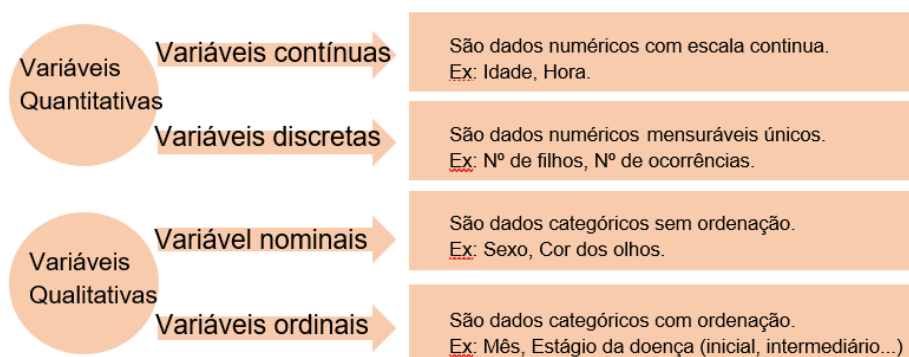
Para aplicação dos algoritmos, além das normalizações necessárias realizado pelo Excel, foi preciso realizar o agrupamento da base de dados no atributo “natureza1\_descricao” por conter 167 instâncias únicas, aplicar a base sem este agrupamento no algoritmo leva muito tempo de processamento talvez até dias. Para

a realização deste agrupamento foi utilizado dois métodos diferentes, o primeiro método com a utilização do Algoritmo de classificação K-Means gerou 5 clusters através da ferramenta R Studio e o segundo agrupamento de forma manual realizado no Excel.

### 3.5 ESTATÍSTICA DESCRITIVA

A estatística descritiva é a ciência que analisa, interpreta e representa conjuntos de dados. Segundo Reis (2002, p. 23) na descritiva primeiramente é identificado o tipo de variável presente na base classificando em dados Qualitativos e quantitativos segundo a Figura 14.

FIGURA 14 - TIPO DE VARIÁVEIS



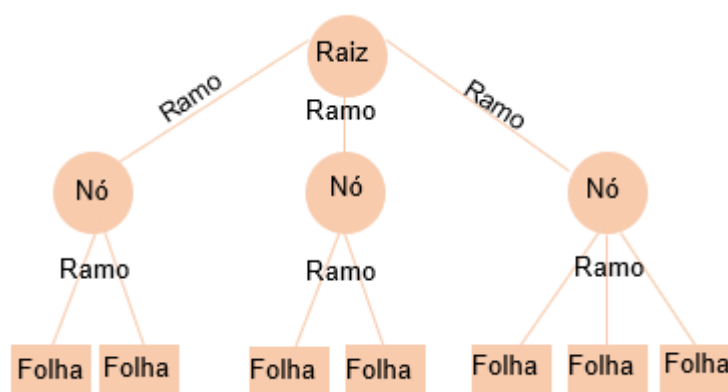
FONTE: Adaptado REIS (2002, p. 23)

A estatística descritiva foi aplicada neste trabalho seguindo o tipo de cada atributo de acordo com a Figura 14, os atributos presentes na base são dados qualitativos nominais e ordinais, portanto, para análise foi realizada uma tabela de frequências, gráficos de barras, mínimo e máximo e a ferramenta utilizada para gerar estas análises no Excel.

### 3.6 MINERAÇÃO DE DADOS

Os algoritmos selecionados para a mineração de dados de acordo com o tipo de dados da base de dados da Guarda Municipal de Curitiba contêm somente dados qualitativos, portanto foi selecionado o algoritmo Árvore de decisão e o Naïve Bayes e ambos foram executados na ferramenta R Studio. A árvore de decisão no R Studio é executada com a função “Rpart, e o resultado se dá através de classificações que resultam em um conjunto de nós de decisão e folhas segundo Pinto, Vieira e Bulhoes (2005). Para exemplificar a árvore de decisão através da Figura 15 é apresentado:

FIGURA 15 - EXEMPLO ÁRVORE DE DECISÃO



FONTE: A Autora (2020)

De acordo com a Figura 15, cada nó da árvore testa um atributo previsor, o primeiro nó na árvore é a raiz e a partir deste nó saem os ramos, cada ramo corresponde a um valor possível do atributo previsor testado e as folhas representam o resultado de todos os testes executados no atributo previsor.

Segundo Mello (2002, p. 26) há muitos anos vem sendo utilizado as redes bayesianas como processo de classificação. O Naïve Bayes é um método que utiliza cálculos de probabilidades para prever a classe mais provável. Para a aplicação deste algoritmo foi realizado através do R Studio através da função “naiveBayes”.

### 3.6 PÓS PROCESSAMENTO

O pós-processamento é a última fase do KDD, e para realização deste capítulo foi considerado o resultado obtido com a estatística descritiva e os algoritmos Árvore

de decisão e o Naïve Bayes. O método utilizado para análise dos algoritmos é através de comparações do resultado da matriz confusão com as taxas de acertos, levando em consideração a comparação do resultado de duas bases sendo uma com dados agrupados manualmente e outra com os dados agrupados pelo K-Means, através deste capítulo foi identificado também as principais tendências e padrões.

#### 4. EXPERIMENTOS E ANÁLISE DOS RESULTADO

No presente capítulo foram realizados os experimentos na base de dados através do pré-processamento de dados, estatística descritiva, mineração de dados e o pós-processamento.

##### 4.1 PRÉ-PROCESSAMENTO

Para se obter uma melhor performance dos algoritmos de mineração de dados o pré-processamento é fundamental para limpeza e tratamento dos dados sendo necessário considerar características importantes da base com a seleção das variáveis mais importantes, exclusão de anomalias, identificação de registros faltantes, normalizações e agrupamentos necessários.

A base de dados da Guarda Municipal ao todo tem 35 atributos e 280.905 registros de ocorrências entre o ano de 2009 até o mês de maio de 2020.

Para a realização da análise deste trabalho foram selecionados os atributos necessários para compor a pesquisa de acordo com a Quadro 11:

QUADRO 11 - ATRIBUTOS SELECIONADOS PARA COMPOR A BASE DE DADOS

(continua)

Atributo	Descrição	Limites de variação ex	Formato	Exemplo
Atendimento_ano	Ano do atendimento da ocorrência	De 2009 até 2019	Int	2009
Atendimento_bairro_nome	Bairro do registro da ocorrência	136 Bairros considerando a Região Metropolitana	varchar	Cabral
Flag_flagrante	Flag para identificar se é um equipamento urbano ou não	Sim ou não	char	Sim

Natureza1_defesa_civil	Flag para identificar se é uma natureza de defesa civil ou não (0 - Não, 1 - Sim)	Sim ou não	bit	Não
Natureza1_descricao	Descrição do tipo de ocorrência	167 Descrições	varchar	Dano
Subcategoria1_descricao	Subcategoria da descrição do tipo de ocorrência	283 Subcategorias	varchar	Vandalismo
Ocorrencia_dia_semana	Dia da semana da ocorrência	De Segunda até Sábado	varchar	Sábado
Ocorrencia_hora	Hora da ocorrência	Períodos (Manhã, Tarde, Noite, Madrugada)	varchar	00:15
Ocorrencia_mes	Mês da ocorrência	12 meses	Int	11
Origem_chamado_descricao	Origem da ocorrência	25 origens	varchar	153
Regional_fato_nome	Regional da ocorrência	9 Regionais em Curitiba e Região Metropolitana	varchar	Matriz
Secretaria_sigla	Secretaria responsável sobre a ocorrência	38 Secretarias	varchar	SMDT

FONTE: SiGesGuarda (2020)

Os 12 atributos presentes no Quadro 11 são os utilizados para a análise deste estudo, os demais 23 atributos não serão utilizados pela pesquisa por repetirem informações, por estarem vazios ou por não serem relevantes para o objetivo da pesquisa conforme o Quadro 12.

QUADRO 12 - ATRIBUTOS RETIRADOS DA ANÁLISE DOS DADOS

(continua)

Atributo	Descrição do seu não uso
Equipamento_urbano_nome	Está incompleto com vários campos vazios
Flag_equipamento_urbano	Não é o foco da pesquisa
Logradouro_nome	Não é o foco da pesquisa e não tem a longitude e latitude
Natureza2_defesa_civi	Além do nome se repetir o campo está vazio
Natureza2_descricao	Além do nome se repetir o campo está vazio
Natureza3_defesa_civil	Além do nome se repetir o campo está vazio
Natureza3_descricao	Além do nome se repetir o campo está vazio
Natureza4_defesa_civil	Além do nome se repetir o campo está vazio
Natureza4_descricao	Além do nome se repetir o campo está vazio
Natureza5_defesa_civil	Além do nome se repetir o campo está vazio
Natureza5_descricao	Além do nome se repetir o campo está vazio
Subcategoria2_descricao	Além do nome se repetir o campo está vazio



Subcategoria3_descricao	Além do nome se repetir o campo está vazio
Subcategoria4_descricao	Além do nome se repetir o campo está vazio
Subcategoria5_descricao	Além do nome se repetir o campo está vazio
Ocorrencia_data	Os valores já se repetem no atributo de mês e hora
Ocorrencia_ano	Os valores se repetem do atributo atendimento_ano
Ocorrencia_codigo	Não é o foco da pesquisa
Operacao_descrica	O campo está vazio
Secretaria_nome	Vai ser utilizado o campo de siglas
5Servico_nome	Não é o foco da pesquisa
Situacao_equipe_descricao	O campo está vazio
Número_protocolo_156	O campo está vazio em várias linhas

FONTE: SiGesGuarda (2020)

No Quadro 12 foi levantado todos os atributos que não serão utilizados na pesquisa. Os atributos (natureza2\_defesa\_civi, natureza2\_descricao, natureza3\_defesa\_civil, natureza3\_descricao, natureza4\_defesa\_civil, natureza4\_descricao, natureza5\_defesa\_civil, natureza5\_descricao, subcategoria2\_descricao, subcategoria3\_descricao, subcategoria4\_descricao, subcategoria5\_descricao) não foram utilizados pois em ambos estão vazios e o atributo repete o nome com outros campos.

Já os atributos (flag Equipamento\_urbano, ocorrencia\_codigo, servico\_nome, logradouro\_nome) não foram utilizados pois não era o foco da pesquisa analisar a rua, o código da ocorrência e o serviço portanto foram retirados da base de dados. Foram retirados os atributos que estavam com a maior parte das instâncias incompletas sendo (Equipamento\_urbano\_nome e Número\_protocolo\_156). O atributo (Ocorrencia\_data, Ocorrencia\_ano e Secretaria\_nome) contém as mesmas informações presentes em outros atributos cujo qual já foram selecionados apresentados no Quadro 11. E por fim foram eliminados os atributos (Operacao\_descrica e Situacao\_equipe\_descricao) por estarem vazios.

#### 4.1.1 Limpeza dos dados

A primeira validação a ser analisada na base de dados são os valores ruidosos. Na primeira análise foram identificados registros de janeiro de 2009 até o mês maio de 2020, porém o objetivo desta pesquisa é a utilização das ocorrências com os ciclos

anuais fechados. Para esta análise foi utilizado o campo Atendimento\_Ano e através de filtros foram selecionados todos que correspondem ao ano de 2020, eliminando da base (15.501) registros e a base de dados ficando com (265.401) ocorrências.

A segunda validação avalia valores ausentes, foram identificados através dos atributos Atendimento\_Ano, Subcategoria1\_Descricao e Subcategoria1\_Descricao ocorrências com dados ausentes.

No atributo Atendimento\_Ano, (14) ocorrências estavam sem valores, porém o campo Ocorrencia\_Ano continha os mesmos atributos, a eliminação nesta validação não foi necessária pois tem valores dependentes e auxiliaram na complementação destes campos, deste modo não perdendo informações para análise.

No atributo Subcategoria1\_Descricao, contém (98.935) ocorrências com dados ausentes, a sua eliminação corre o risco de se perder valiosas informações, a opção selecionada para esta pesquisa foi a verificação através de uma tabela cruzada com o campo Subcategoria1\_Descricao e Natureza1\_Descricao verificando se em todos os campos de natureza são vazios na subcategoria, os que forem vazios foi repetido o nome da natureza para o campo subcategoria, os que não forem vazios da natureza com a subcategoria é retirado da análise por não se ter o valor correto. A Tabela 3 se refere as ocorrências que tem subcategoria e valores ausentes.

TABELA 3 - RELAÇÃO DE VALORES AUSENTES NO ATRIBUTO NATUREZA DESCRIÇÃO E SUBCATEGORIA DESCRIÇÃO

NATUREZA1_DESCRICAÇÃO	subcategoria1_descricao (Vazias)	subcategoria1_descricao (Total Geral)
Achado		610
Alarmes	7	10.032
Animais	1	13.028
Apoio	11	40.474
Câmeras de videomonitoramento	1	5
Dano	9	33.354
Disparo de arma	257	1.081
Encaminhamento	5	1.573
Furto	39	6.047
Invasão	22	11.477
Lesão Corporal	29	793
Órgãos acionados		7
Porte Ilegal		556
Risco de acidente / à vida	4	1.207
Roubo	70	16.087

Substância Ilícita	5	16.443
Substância Lícita		326
Tentativa		2.310
Trânsito	1	8.519
Veículo	1	2.708

FONTE: SiGesGuarda (2020)

Foram eliminados todos os valores vazios da tabela sendo 462 ocorrências de acordo com a Tabela 3. O restante das ocorrências vazias do atributo Natureza1\_Descricao não tem nenhuma ocorrência registrada com Subcategoria1\_Descricao, portanto para o atributo Subcategoria1\_Descricao foi preenchido com o valor da Natureza1\_Descricao não perdendo (98.935) ocorrências da pesquisa.

Através das validações de ausência de dados na base foram identificados dois atributos com dados vazios, um é o atributo Atendimento\_Ano à qual não precisou eliminar nenhuma ocorrência, já o segundo atributo é a Subcategoria1\_Descricao foi eliminado (462) registros, totalizando a base com (264.939) ocorrências.

A terceira validação identificado na base de dados são os valores inconsistentes, foram realizadas quatro análises. A primeira análise foi aplicada no atributo atendimento\_bairro\_nome à qual identificou 4 situações diferente, a primeira situação foi identificada (9) ocorrências registradas como “bairro não informado” sendo eliminados da base de dados.

A segunda análise encontrada foi no atributo atendimento\_bairro\_nome, (3) ocorrências estavam registradas como “bairro não localizado” eliminando os mesmos da pesquisa.

A terceira análise foi identificada no atributo Regional\_Fato\_Nome na base cotinha a regional Tatuquara, porem Tatuquara é um bairro pertencente a regional Pinherinho. Na Regional Tatuquara continha (3620) ocorrências distribuídas de acordo com a Tabela 4.

TABELA 4 - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL TATUQUARA

Bairro	Quantidade de Dados
BACACHERI	1
CAMPO COMPRIDO	1
CAMPO DE SANTANA	1.026

CAPÃO RASO	1
CAXIMBA	153
CIDADE INDUSTRIAL	1
PORTÃO	1
TATUQUARA	2.435
UMBARÁ	1

FONTE: A autora (2020)

Através da Tabela 4 percebe-se que existem muitos bairros dentro da Regional Tatuquara (1185) ocorrências pertencentes aos bairros Bacacheri, Campo Comprido, Campo de Santana, Capão Raso, Caximba, Cidade industrial, Portão e Umbara foram eliminados da pesquisa. Já o Bairro cujo nome está Tatuquara foi corrigido, com sua Regional passando a ser Pinheirinho, que contém (2.435) registros.

Já a quarta e última análise dos valores inconsistentes foi identificado no atributo Atendimento\_bairro\_nome, foi realizada através de uma tabela cruzada com o atributo Regional\_fato\_nome e identificado vários bairros com dados em regionais erradas, para esta análise foi utilizado como referência o mapa das regionais e bairros de Curitiba disponibilizados pelo site<sup>2</sup> da FAS (Fundação de Ação Social) de Curitiba e através da tabela dinâmica analisado os dados com inconsistências e sem inconsistências. Através dos Quadros 13 e 14 contém a classificação de bairros nas regionais segundo FAS.

QUADRO 13 - RELAÇÃO DE BAIRROS NAS REGIONAIS 1

BAIRRO NOVO	BOA VISTA	BOQUEIRAO	CAJURU	CIC
GANCHINHO	ABRANCHES	ALTO BOQUEIRAO	CAJURU	AUGUSTA
SÍTIO CERCADO	ATUBA	BOQUEIRAO	CAPÃO DA IMBUIA	CIDADE INDUSTRIAL
UMBARÁ	BACACHERI	HAUER	GUABIROTUBA	RIVIERA
	BAIRRO ALTO	XAXIM	JARDIM DAS AMÉRICAS	SÃO MIGUEL
	BARREIRINHA		UBERABA	
	BOA VISTA			
	CACHOEIRA			
	PILARZINHO			
	SANTA CÂNDIDA			
	SÃO LOURENÇO			
	TABOÃO			
	TARUMÃ			
	TINGUI			

<sup>2</sup> Fonte: <https://fas.curitiba.pr.gov.br/maparegional.aspx>

FONTE: FAS (2020)

Através do Quadro 13 observa-se a distribuição de bairros pelas regionais Bairro novo, Boa Vista, Boqueirão, Cajuru e CIC.

QUADRO 14 - RELAÇÃO DE BAIRROS NAS REGIONAIS 2

MATRIZ	PINHEIRINHO	PORTAO	SANTA FELICIDADE
AHÚ	CAMPO DE SANTANA	ÁGUA VERDE	BUTIATUVINHA
ALTO DA GLÓRIA	CAPÃO RASO	FANNY	CAMPINA DO SIQUEIRA
ALTO DA RUA XV	CAXIMBA	FAZENDINHA	CAMPO COMPRIDO
BATEL	PINHEIRINHO	GUAÍRA	CASCATINHA
BIGORRILHO	TATUQUARA	LINDÓIA	LAMENHA PEQUENA
BOM RETIRO		NOVO MUNDO	MOSSUNGUÊ
CABRAL		PAROLIN	ORLEANS
CENTRO		PORTAO	SANTA FELICIDADE
CENTRO CÍVICO		SANTA QUITÉRIA	SANTO INÁCIO
CRISTO REI		VILA IZABEL	SÃO BRAZ
HUGO LANGE			SÃO JOÃO
JARDIM BOTÂNICO			SEMINÁRIO
JARDIM SOCIAL			VISTA ALEGRE
JUVEVÊ			
MERCÊS			
PRADO VELHO			
REBOUÇAS			
SÃO FRANCISCO			

FONTE: FAS (2020)

Já através do Quadro 14 observa-se a distribuição de bairros pelas regionais Matriz, Pinheirinho, Portão e Santa Felicidade.

Para análise das relações de bairros dentro das regionais foi seguido a referência dos Quadros 13 e 14, as inconsistências identificadas de cada regional está apresentado em forma de Tabela. Através da Tabela 5 observa-se a distribuição de bairros dentro da regional Bairro Novo, os demais bairros analisados estão disponíveis no apêndice B até o J.

TABELA 5 - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BAIRRO NOVO

Bairro	Total de Ocorrências	Eliminado?
ALTO BOQUEIRÃO	4	SIM
CAJURU	1	SIM
CAXIMBA	1	SIM
GANCHINHO	1.084	NÃO
PINHEIRINHO	4	SIM
SÍTIO CERCADO	16.061	NÃO
UBERABA	1	SIM
UMBARÁ	1.260	NÃO
XAXIM	1	SIM
TOTAL DE ERROS	12	
<b>TOTAL SEM ERROS</b>	<b>18.404</b>	

FONTE: A autora (2020)

Dentro da redistribuição da regional e bairros na tabela dinâmica apresentada na Tabela 5 percebe-se que a regional Bairro novo na base continha seis bairros a mais a qual não pertenciam àquela regional sendo somente os bairros (Ganchinho, Sítio Cercado e Umbará), foram eliminados os bairros não pertencentes da base através da identificação desta inconsistência e a regional ficou com (18.404) ocorrências registradas. A análise de inconsistências encontradas na Região metropolitana continham (19) bairros que não pertenciam, e (60) ocorrências foram eliminadas da base através da identificação desta inconsistência e o a Região Metropolitana ficando com (63) bairros.

A partir da análise da Tabela 5 e as Tabelas no apêndice B até o J sobre a relação do atributo Regional\_fato\_nome e Atendimento\_Bairro\_Nome foi elaborado a a Tabela 6 a qual contém um resumo com todas os dados das regionais que tiveram inconsistências (Erros) e sem inconsistências.

TABELA 6 - RELAÇÃO DE INCONSISTÊNCIAS ENTRE AS REGIONAIS

Regional	Dados com inconsistências	Dados sem inconsistências
BAIRRO NOVO	12	18.405
BOA VISTA	21	28.609

BOQUEIRÃO	10	25.516
CAJURU	59	23.238
CIC	51	19.484
MATRIZ	71	81.038
PORTÃO	288	25.728
PINHEIRINHO	1.645	18.614
SANTA FELICIDADE	42	20.788
REGIÃO METROPOLITANA	60	73
<b>TOTAL</b>	<b>2.249</b>	<b>261.493</b>

FONTE: A Autora (2020)

Através da Tabela 6 observa-se que as regionais pinheirinho e portão foram as que mais continham inconsistências, ao todo somando todas as inconsistências das regionais foram eliminadas (2249) ocorrências.

Através da validação de inconsistência foram identificadas muitas ocorrências com erros, e decidiu-se pela sua eliminação para não comprometer o resultado. Os dados foram eliminados da base através de 4 tipos de situações. A primeira situação foi identificada (9) ocorrências registradas como bairro não informado, a segunda situação (3) ocorrências estavam registradas como “bairro não localizado”, a terceira situação continha muitos bairros dentro de uma regional que não existe a Tatuquara (1185) ocorrências foram eliminadas pertencente aos bairros “*Bacacheri, Campo Comprido, Campo de Santana, Capão Raso, Caximba, Cidade industrial, Portão e Umbara*” (2.435) registros cujo nome estava Tatuquara foram corrigidos passando para a Regional Pinheirinho. Foram eliminadas (2.249) ocorrências apresentadas na tabela 6 mais (1.197) ocorrências apresentadas neste parágrafo totalizando (3.446) ocorrências eliminadas.

Não foram identificados valores redundantes na base de dados, apenas as três validações apresentadas anteriormente.

Através desta sessão de limpeza dos dados e análise das quatro validações foram eliminadas (19.409) ocorrências e a base de dados ficando com (261.497) ocorrências, redistribuídas conforme a Tabela 7:

TABELA 7 - RELAÇÃO DE DADOS ELIMINADOS APLICADOS EM QUATRO VALIDAÇÕES

<b>Tipos de validações</b>	<b>Valores retirados</b>
<b>Valores Ruidosos</b>	15.501
<b>Valores Ausentes</b>	462
<b>Valores Inconsistentes</b>	3.446
<b>Valores Redundantes</b>	0
<b>TOTAL</b>	19.409

FONTE: A Autora (2020)

Observa-se que a maior concentração dos registros eliminados está na validação de valores ruidosos sendo os períodos que não serão utilizados e na sequência valores inconsistentes.

#### 4.1.2 Tratamento da base de dados

O tratamento da base de dados neste trabalho se dá através da normalização, categorização e classificação dos dados. A normalização dos dados auxiliara na análise realizada no R Studio, deixando a base de dados pronta para ser realizada as estatísticas descritivas e a aplicação dos algoritmos de mineração de dados.

O primeiro tratamento realizado é nos atributos nominais com a retirada dos acentos apresentados no Quadro 15.

QUADRO 15 - INSTÂNCIAS NOMINAIS NORMALIZADAS

<b>Instancias</b>	<b>Instancias corrigidas</b>
BOQUEIRÃO	BOQUEIRAO
PORTÃO	PORTAO
REGIÃO METROPOLITANA	REGIAO_METROPOLITANA
SÁBADO	SABADO
NÃO	NAO

FONTE: A Autora (2020)



De acordo com o Quadro 15 foram normalizadas as instancias dos atributos (Regional\_fato\_nome, Ocorrencia\_dia\_semana e Flag\_flagrante).

O segundo tratamento realizado na base de dados são os acréscimos de “Underline” em todas as instâncias com espaço para serem reconhecidas como únicas pelo R Studio.

O atributo Natureza1\_defesa\_civil tem os dados no formato de 0 e 1 de acordo com a Tabela 8 foi normalizado as instâncias binarias para nominal.

TABELA 8 - NORMALIZAÇÃO DAS INSTANCIAS DO ATRIBUTO NATUREZA1\_DEFESA\_CIVIL

Instancias	Nova Instancia	Total de ocorrências
0	NAO	253.109
1	SIM	8.388

FONTE: A Autora (2020)

A normalização do atributo Natureza1\_defesa\_civil transforma os valores numéricos para categóricos, sendo iguais a “0” para NAO e 1 para SIM conforme a Tabela 8.

O quarto tratamento realizado é a conversão dos anos que estão em formato numérico no atributo atendimento\_ano para categóricos como fatores no R Studio, para que possa ser realizada a análise. A categorização ficou com os anos de 2009 até 2019 seguindo o alfabeto de A até K, sendo A 2009 e K 2019.

O quinto tratamento realizado é no atributo Origem\_chamado\_descricao, conforme observa-se no Quadro 16.

QUADRO 16 - CATEGORIZAÇÃO DE VARIÁVEIS NO ATRIBUTO ORIGEM\_CHAMADO\_DESCRICAO

Instancias	Nova Instancia
153	GUARDA_MUNICIPAL
156	CENTRAL_DE_ATENDIMENTO_PREFEITURA
199	DEFESA_CIVIL
CIOSP_ (190)	CIOSP

FONTE: A Autora (2020)

Foram transformados os valores que são numéricos dentro do atributo para categórico no atributo (Origem\_chamado\_descricao) conforme o Quadro 16.

O sexto tratamento realizado na base de dados ocorreu no atributo Ocorrencia\_hora classificando as horas conforme o Quadro 17:

QUADRO 17 - DISTRIBUIÇÃO DE HORAS POR PERÍODO

Período	Métricas
Madrugada	00:01 até as 6:00
Manhã	06:01 até as 12:00
Tarde	12:01 até as 18:00
Noite	18:01 até as 00:00

FONTE: A Autora (2020)

Foi realizado o agrupamento do atributo Ocorrencia\_hora considerando períodos sendo (Madrugada, Manhã, Tarde e Noite) seguindo as métricas apresentadas no Quadro 17.

O sétimo tratamento na base de dados é no atributo Ocorrencia\_mes conforme o Quadro 18.

QUADRO 18 - CATEGORIZAÇÃO DO ATRIBUTO OCORRÊNCIA MÊS

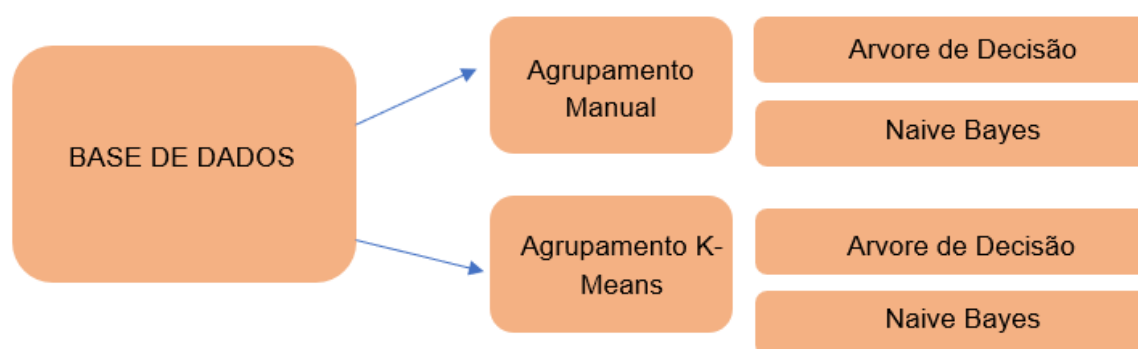
Mês Numérico	Mês categórico
1	JANEIRO
2	FEVEREIRO
3	MARÇO
4	ABRIL
5	MAIO
6	JUNHO
7	JULHO
8	AGOSTO
9	SETEMBRO
10	OUTUBRO
11	NOVEMBRO
12	DEZEMBRO

FONTE: A Autora (2020)

De acordo com o Quadro 18, foi categorizado o atributo Ocorrencia\_mes convertendo as instâncias numéricas para categóricas.

O último tratamento foi no atributo Natureza1\_descricao que contém 167 instancias únicas, e para aplicação dos algoritmos de mineração dados é necessário realizar agrupamentos para que o processamento seja realizado no R Studio e não leve horas talvez até dias. Para a realização deste agrupamento foi considerado dois métodos diferentes, um método com a utilização do Algoritmo de classificação K-Means e o segundo manual de acordo com a Figura 16.

FIGURA 16 - AGRUPAMENTO DO ATRIBUTO NATUREZA\_DESCRIÇÃO



FONTE: A Autora (2020)

Após a realização dos agrupamentos considerando a base de dados gerada manualmente agrupada pelo K-Means foram aplicados os algoritmos duas vezes conforme descrito na Figura 16.

Para a realização do agrupamento manual primeiramente foi verificado todas as instâncias únicas no atributo natureza1\_descricao e classificado em 5 grupos e distribuído as descrições conforme a Tabela 9.

TABELA 9 - CLASSIFICAÇÃO DOS GRUPOS BASE MANUAL

Agrupamento Manual	Total de descrição únicas por grupo	Grupo
Grupo_1	23	Natureza
Grupo_2	36	Apoio

Grupo_3	27	violência
Grupo_4	35	Dano
Grupo_5	46	Outros

FONTE: A Autora (2020)

Através do apêndice K é apresentado de forma detalhada como ficaram divididas as descrições dentro de cada grupo através do agrupamento manual conforme apresentado as divisões na Tabela 9.

O segundo agrupamento realizado é o K-Means, no R Studio conforme a Figura 17.

FIGURA 17 - SCRIPT DE AGRUPAMENTO K-MEANS

```
##Instalando pacotes
install.packages("fastDummies")
install.packages("rlang")

## Carregando os dados
dados <- read.csv2("Guarda_Municipal.csv")

#Traforma as instacias que são categoricas para binarias
grupos <- fastDummies::dummy_cols(dados$NATUREZA1_DESCRICA0)
grupos <- grupos[, -1] # remove a primeira coluna de texto
d <- dist(t(grupos)) # t transpõe a matriz

#Agrupamento e classifica através do K-means
km <- kmeans(d, 5) # definir o numero de grupos
table(km$cluster)

#Viauzaliza a tabela
view(km$cluster)

#gera um novo CSV
write.csv2(km$cluster, "kmeans.csv")
```

FONTE: A Autora (2020)

Pode-se observar na Figura 17 que primeiramente foram instalados os pacotes necessários para a classificação, em sequência foi carregada a base de dados, foram transformadas as instâncias categóricas em binárias, e por fim foi executado o k-Means com a definição de 5 *clusters* (grupos), o resultado gerado foi transformado em um novo arquivo CSV com os agrupamentos já realizados.

Após a realização do agrupamento pelo K-mens foi analisada e desenvolvida a Tabela 10 a qual contém a distribuição das instâncias únicas do atributo *natureza1\_descricao* por grupos.

TABELA 10 - CLASSIFICAÇÃO DOS GRUPELO PELO K-MEANS

Agrupamento no R	Total de descrição únicas por grupo	Descrição do grupo
Grupo_1	106	Descrições Aleatórias
Grupo_2	28	Descrições Aleatórias
Grupo_3	19	Descrições Aleatórias
Grupo_4	6	Descrições Aleatórias
Grupo_5	8	Descrições Aleatórias

FONTE: A Autora (2020)

Ao serem gerados os grupos no K-Means, estes ficaram com as descrições de forma aleatória. Por exemplo no grupo\_5 da Tabela 10 contém a descrição Agressão\_física/verbale e animais. Através do apêndice L contém de forma detalhadamente como ficaram divididas as descrições dentro de cada grupo.

## 4.2 ESTATÍSTICA DESCRITIVA

Através da identificação dos tipos de variáveis presentes na base de dados, definiu-se o tipo de análise realizada nesta seção. Por ser uma base com dados qualitativos nominal e ordinal conforme apresentado no Quadro 19 as análises realizadas são através da tabela de frequências, gráficos de barras, mínimo e máximo.

QUADRO 19 - CLASSIFICAÇÃO DOS ATRIBUTOS

Atributo	Tipo	Variação
Atendimento_ano	Variável Qualitativa Ordinal	De 2009 até 2019 (A até K)
Origem_chamado_descricao	Variável Qualitativa Nominal	25 origens
Secretaria_sigla	Variável Qualitativa Nominal	38 Secretarias
Flag_flagrante	Variável Qualitativa Nominal	Sim ou não
Natureza1_defesa_civil	Variável Qualitativa Nominal	Sim ou não
Natureza1_descricao	Variável Qualitativa Nominal	167 Descrições
Subcategoria1_descricao	Variável Qualitativa Nominal	283 Subcategorias
Ocorrencia_mes	Variável Qualitativa Ordinal	12 meses
Ocorrencia_dia_semana	Variável Qualitativa Ordinal	De Segunda até Sábado
Ocorrencia_hora	Variável Qualitativa Ordinal	Períodos (Manhã, Tarde, Noite, Madrugada)
Regional_fato_nome	Variável Qualitativa Nominal	9 Regionais em Curitiba e Região Metropolitana

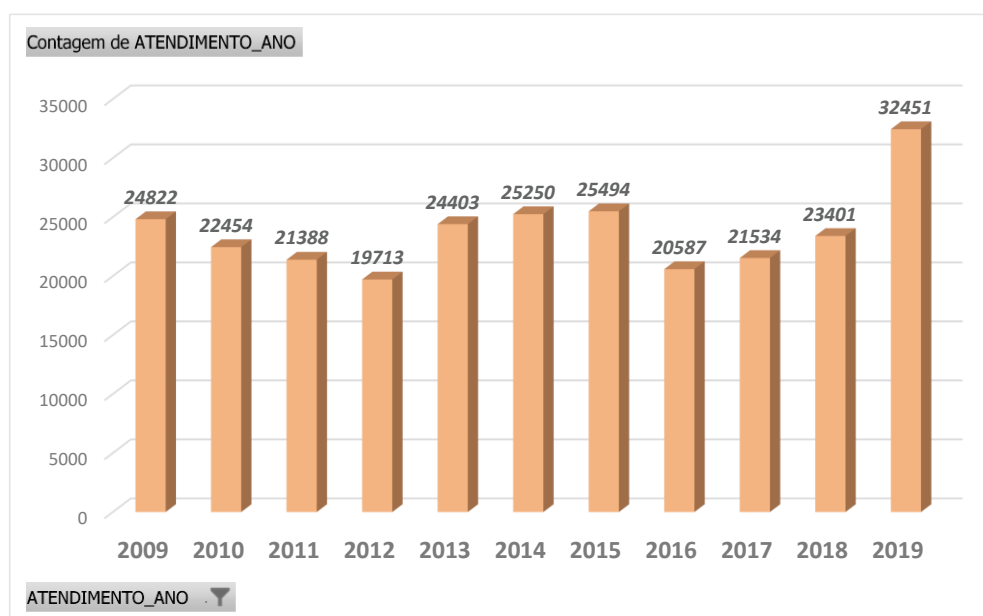
Atendimento_bairro_nome	Variável Qualitativa Nominal	136 Bairros considerando a Região Metropolitana
Codigodescricao	Variável Qualitativa Nominal	5 Grupos

FONTE: A Autora (2020)

Através do Quadro 19 observa-se que a maior concentração dos tipos de variável é nominal, sendo 9 atributos nominal e 4 ordinais.

O primeiro atributo analisado é o Atendimento\_ano, e através do Gráfico 1 é possível verificar os anos que mais tiveram ocorrências e os que menos tiveram.

GRÁFICO 1 - GRÁFICO DE DISTRIBUIÇÃO DE OCORRÊNCIAS POR ANO



FONTE: A Autora (2020)

Através do Gráfico 1 observou-se que o ano que mais teve ocorrências registradas foi o de 2019 com um total de (32.451) registros, 10 mil a mais em comparação aos outros anos, já o ano que menos teve registros foi em 2012 com (19.713) ocorrências e o restante dos anos teve uma média de (23.777).

Para a segunda análise no atributo origem\_chamado\_descricao, foi relacionado com Atendimento\_Ano e observou-se que a base contém (25) origens de ocorrências. Na Tabela 11 contém o ranking de 10 origens que mais aparecem na base, as demais 25 origens estão colocadas no apêndice M.

TABELA 11 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE ORIGENS DE OCORRÊNCIAS POR ANO

Origem	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
153	7.135	7.944	6.498	7.934	11.280	11.502	12.428	8.881	8.743	9.235	9.706	101.286
IMEDIATA	3.338	3.163	3.867	4.186	4.197	3.473	3.351	4.284	6.280	7.194	14.945	58.278
CIOSP (190)	4.718	2.769	1.482	1.601	1.631	2.007	2.082	1.862	1.639	1.371	827	21.989
AOS GMS	1.824	1.670	2.172	1.596	2.019	2.450	1.983	1.587	1.288	1.230	2.092	19.911
156	2.365	1.893	1.647	1.208	1.792	1.733	924	302	119	612	884	13.479
À VIATURA	1.179	805	955	866	1.123	1.267	1.859	1.126	1.069	982	1.600	12.831
FONE	887	809	740	846	899	1.078	1.103	972	803	1.009	570	9.716
G5	355	373	529	382	712	867	791	755	743	516	330	6.353
SIGA	2.308	1.786	1.503	210	5	11	13	2	2	69	15	5.924
OFÍCIO	106	528	1.341	361	202	261	257	174	160	140	73	3.603
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

FONTE: A Autora (2020)

Na Tabela 11 o principal meio de origem das ocorrências na base de dados é o 153 (Telefone da Guarda Municipal de Curitiba) com um total de (101.286) ocorrências somadas entre os anos, destacando o ano que mais se teve registros em 2015 com (12.428) ocorrências, o período que menos teve ocorrências nesta origem foi em 2011 com (6.498) registros. Já a origem (OFÍCIO) nesta relação dos 10 principais meios de origem de ocorrências foi o que menos teve dados, com um total de (3.603) ocorrências, o ano que mais se teve ocorrências nesta origem foi em 2011 com (1.341) registros e o ano que menos teve ocorrências foi em 2019 com (73) variáveis.

Com o cruzamento dos atributos secretaria\_sigla e Atendimento\_Ano obteve-se uma tabela de frequências absolutas e observou-se que a mesma contém (38) secretarias registradas entre as ocorrências. Para esta análise foram selecionadas as (10) secretarias que mais contêm ocorrências registradas entre as (38) segundo a Tabela 12, no apêndice N contém a tabela completa com as demais secretarias.

TABELA 12 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE SECRETARIAS/ANO

Secretaria	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
SMDT	8.460	9.341	12.898	9.758	12.262	14.011	11.288	10.715	11.541	12.819	24.625	137.718
SMMA	7.011	5.966	2.354	4.057	5.821	5.321	5.527	3.347	3.336	3.127	2.839	48.706
URBS	3.094	2.171	2.153	2.654	2.944	2.990	4.486	2.835	2.554	2.362	1.458	29.701

SMS	1.688	998	1.152	1.287	1.458	1.163	1.713	1.491	1.471	1.591	1.204	15.216
SME	1.107	1.136	1.098	789	930	957	1.303	1.212	1.241	1.229	848	11.850
FAS	2.818	2.288	1.350	662	504	271	416	336	458	618	487	10.208
SMU	315	178	63	201	177	212	365	244	375	299	217	2.646
SETRAN				25	47	50	53	52	21	792	167	1.207
SMSAN	78	97	106	101	63	72	75	81	81	71	47	872
SMEM								80	189	202	193	664
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

FONTE: A Autora (2020)

Na Tabela 12, a secretaria “Superintendência de Trânsito - Secretaria Municipal de Defesa Social e Trânsito (SMDT)” contém mais ocorrências registradas que as demais, com um total de (137.718) ocorrências somadas entre os períodos, concentrando a maior parte destes valores no ano de 2019 com (24.625) dados e o período que menos teve registros nesta secretaria foi em 2009 com (8.460) variáveis. Já a secretaria “Secretaria Municipal Extraordinária da Mulher (SMEM)”, nesta análise das 10 principais secretarias que aparecem na base de dados é a que contém menos ocorrências (664) registros somados entre os anos. Essa secretaria, por ser nova, contém registros cadastrados desde 2016 e percebe-se uma evolução entre os anos aumentando cada vez mais o número de ocorrências registradas.

A quarta análise realizada é no atributo Flag\_flagrante, através de uma tabela cruzada de frequências absolutas cruzando os atributos Flag\_Equipamento\_Urbano e Atendimento\_Ano apresentada na Tabela 13.

TABELA 13 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE FLAGRANTES POR ANOS

Ano	NÃO	SIM	Total Geral
2009	23.939	883	24.822
2010	21.504	950	22.454
2011	20.458	930	21.388
2012	18.830	883	19.713
2013	23.058	1.345	24.403
2014	23.898	1.352	25.250
2015	24.191	1.303	25.494
2016	19.528	1.059	20.587
2017	20.492	1.042	21.534
2018	22.237	1.164	23.401



2019	30.579	1.872	32.451
<b>Total Geral</b>	<b>248.714</b>	<b>12.783</b>	<b>261.497</b>

FONTE: A Autora (2020)

Através da Tabela 13 percebe-se que teve menos pego em flagrante do que não pego em flagrante, (248.714) ocorrências não tiveram flagrante destacando o ano de 2019 que mais teve ocorrências com (30.579) registros e o ano de 2012 que menos teve ocorrências com (18.830) registros. Já as ocorrências que tiveram flagrante foram as que menos tiveram ocorrências, o ano de 2019 teve mais ocorrências com pegos em flagrante em um total de (1.872) registros já o ano de 2009 e 2012 teve menos ocorrências (883) em registros de cada uma.

A quinta análise na base de dados se deu através da relação do atributo Natureza1\_defesa\_Civil e Atendimento\_Ano, a análise apresenta as ocorrências cujo atendimento foi de natureza civil ou não, relacionado aos anos conforme apresentado na Tabela 14.

TABELA 14 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE DEFESA CIVIL POR ANOS

Ano	NÃO	SIM	Total Geral
2009	23.720	1.102	24.822
2010	21.648	806	22.454
2011	20.193	1.195	21.388
2012	18.852	861	19.713
2013	23.588	815	24.403
2014	24.804	446	25.250
2015	25.024	470	25.494
2016	20.269	318	20.587
2017	21.269	265	21.534
2018	22.888	513	23.401
2019	30.854	1.597	32.451
<b>Total Geral</b>	<b>253.109</b>	<b>8.388</b>	<b>261.497</b>

FONTE: A Autora (2020)

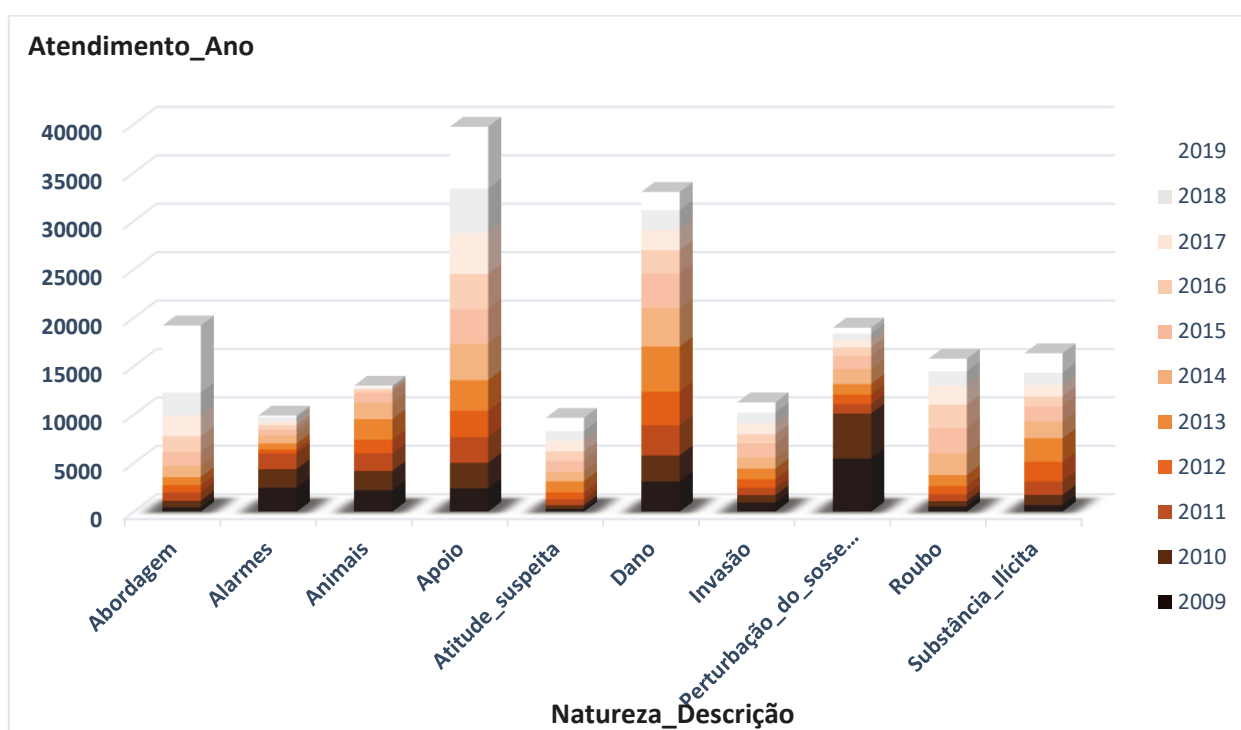
É possível observar na Tabela 14 que a maior parte das ocorrências registradas não são de natureza civil sendo o total de (253.109) ocorrências, dentre as ocorrências (30.854) foram registradas no ano de 2019 sendo o ano com mais ocorrências, já o

ano de 2012 foi o ano que menos registrou ocorrências com (18.852) registros. As ocorrências que são de natureza Civil somaram (8.388) registros entre eles (1.597) ocorrências foram geradas em 2019 e no ano de 2017 conteve menos ocorrências (265) no total.

O atributo Natureza1\_descricao contém 167 tipos de descrições únicas, para a aplicação do algoritmo de mineração foram realizados dois agrupamentos conforme explicado no capítulo anterior de tratamento nas bases de dados. Neste capítulo de estatística foram analisados os dados sem agrupamento e com agrupamento tanto manual e k-Means.

A análise do atributo Natureza1\_descricao sem o agrupamento foi considerado o ranking das 10 principais descrições presentes na base de dados, cruzando o atributo Atendimento\_Ano conforme apresentado no Gráfico 2.

GRÁFICO 2 - GRÁFICO DA DISTRIBUIÇÃO DAS 10 PRINCIPAIS DESCRIÇÕES POR ANO



FONTE: A Autora (2020)

Através do Gráfico 2 é possível observar que as descrições que mais tiveram ocorrências foram apoio e dano, neste ranking de 10 principais descrições. Na descrição perturbação da barra de cor marrom escuro (2009) e marrom claro (2010) está em destaque por apresentar um maior número de ocorrências para este tipo de

descrição. A Tabela 15 contém de forma detalhada as 10 principais descrições ordenados de forma decrescente a quantidade de ocorrências registradas, as demais 167 descrições estão detalhadas no apêndice O.

TABELA 15 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE DESCRIÇÕES POR ANOS

Descrição	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
Abordagem	433	678	865	747	832	1.169	1.467	1.583	2.161	2.354	6.927	19.216
Alarmes	2.449	1.927	1.616	410	627	840	602	467	394	402	171	9.905
Animais	2.198	1.988	1.823	1.399	2.129	1.721	1.031	331	114	97	162	12.993
Apoio	2.397	2.641	2.656	2.708	3.163	3.769	3.567	3.629	4.232	4.592	6.379	39.733
Atitude suspeita	273	384	614	699	1.148	991	1.120	978	1.096	1.014	1.355	9.672
Dano	3.093	2.699	3.116	3.448	4.680	3.966	3.600	2.392	2.146	1.990	1.875	33.005
Invasão	924	776	750	885	1.108	1.155	1.458	915	1.151	1.088	1.061	11.271
Perturbação do sossego	5.481	4.637	983	973	1.092	1.561	1.327	881	785	668	602	18.990
Roubo	531	537	699	862	1.143	2.222	2.655	2.375	1.965	1.497	1.302	15.788
Substância ilícita	652	1.064	1.393	2.038	2.427	1.719	1.589	1.002	1.179	1.271	2.012	16.346
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>26.1497</b>

FONTE: A Autora (2020)

Através das análises apresentadas na Tabela 15 percebe-se que a descrição “Apoio” foi a que mais apresentou ocorrências com um total de (39.733) registros somados entre os anos, destas ocorrências no ano de 2019 foi o que mais teve registros com (6.379), já o ano de 2009 teve menos ao todo com (2.397) ocorrências. A descrição “Atitude Suspeita” foi a que menos teve ocorrências no ranking das 10 descrições com mais ocorrências, ele registrou (9.672) ocorrências no total, destas (1355) foram registradas no ano que mais se teve em 2019, já o ano de 2009 foi o que menos teve ocorrências nesta descrição (273) ocorrências.

A segunda análise no atributo Natureza1\_descricao foi realizada na base de dados com o agrupamento manual, segundo a Tabela 16 percebe-se como ficou distribuído os grupos:

TABELA 16 - REDISTRIBUIÇÃO DA CLASSIFICAÇÃO DOS GRUPELO PELO MANUAL

<b>Agrupamento Manual</b>	<b>Total de descrição únicas por grupo</b>	<b>Total de ocorrências por grupo</b>	<b>Grupo</b>
Grupo_1	23	22.791	Natureza
Grupo_2	36	60.466	Apoio
Grupo_3	27	68.353	violência
Grupo_4	35	65.074	Dano
Grupo_5	46	44.813	Outros

FONTE: A Autora (2020)

De acordo com a Tabela 16 percebe-se que o grupo que mais tem descrições únicas é o grupo\_5 “outros”, porém apesar de ter a maior concentração não é o que mais tem ocorrências. O grupo que mais contém ocorrências é o grupo\_3 “violência” com (68353) ocorrências e 27 descrições únicas.

A terceira análise realizada no atributo Natureza1\_descricao foi realizada na base de dados com o agrupamento pelo k-Means com as distribuições dos grupos conforme a Tabela 17.

TABELA 17 - REDISTRIBUIÇÃO DA CLASSIFICAÇÃO DOS GRUPOS PELO K-MEANS

<b>Agrupamento no R</b>	<b>Total de descrição únicas por grupo</b>	<b>Total de ocorrências por grupo</b>	<b>Descrição do grupo</b>
Grupo_1	106	3.907	Descrições Aleatórias
Grupo_2	28	11.435	Descrições Aleatórias
Grupo_3	19	33.765	Descrições Aleatórias
Grupo_4	6	14.3078	Descrições Aleatórias
Grupo_5	8	69.312	Descrições Aleatórias

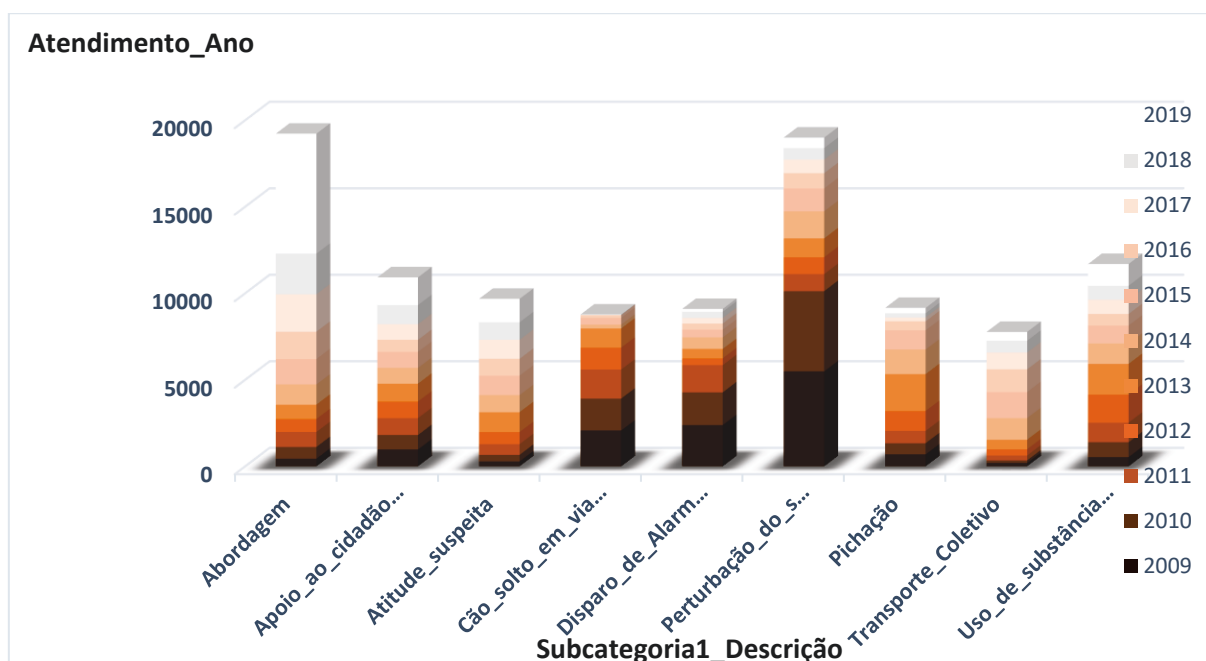
FONTE: A Autora (2020)

De acordo com a Tabela 17 nota-se que apesar do grupo\_1 possuir mais descrições únicas é o grupo que menos tem ocorrências, existem 106 descrições e

3907 ocorrências. O grupo 4 concentra a maior parte das ocorrências da base de dados com 143078 e somente 6 descrições logo percebe-se que estas 6 descrições são as mais utilizadas na base de dados da Guarda Municipal de Curitiba.

Através do cruzamento do atributo Subcategoria1\_descricao e Atendimento\_Ano foi realizada a sétima análise no capítulo, o atributo Subcategoria contém 283 tipos e para a análise foram utilizadas as 10 principais subcategorias apresentadas conforme o Gráfico 3.

GRÁFICO 3 - GRÁFICO DE DISTRIBUIÇÃO DAS 10 PRINCIPAIS SUBCATEGORIAS1\_DESCRÇÃO



FONTE: A Autora (2020)

Através do Gráfico 3 é possível observar que as duas principais subcategorias utilizadas nas ocorrências é abordagem e perturbação o sossego. Na subcategoria perturbação de sossego a maior concentração está em 2009 e 2010. Para melhor análise, a Tabela 18 contém de forma detalhada cada subcategoria distribuída entre os anos ordenados no ranking de 10 principais subcategorias e o apêndice R contém a análise completa.

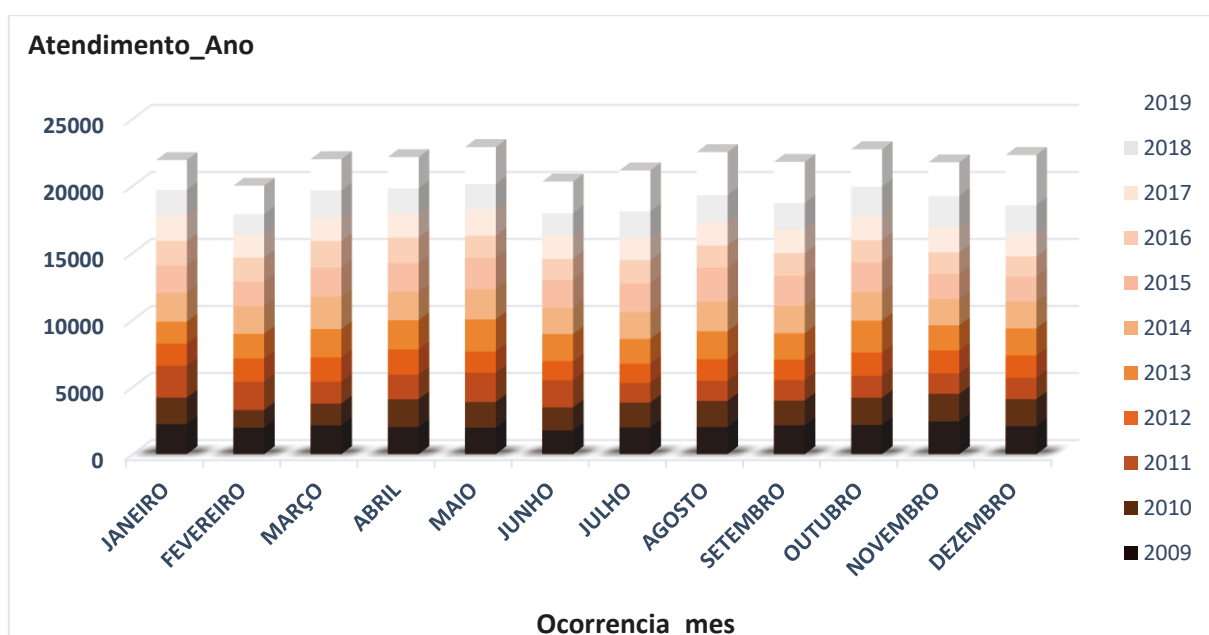
TABELA 18 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE SUBCATEGORIAS POR ANOS

Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
Abordagem	433	678	865	747	832	1.169	1.467	1.583	2.161	2.354	6.927	19.216
Perturbação do sossego	5.481	4.637	983	973	1092	1.561	1.327	881	785	668	602	18.990
Uso de substância ilícita	525	862	1.127	1.620	1.777	1.180	1.036	669	828	807	1253	11.684
Apoio ao cidadão - prestação de socorro/sa- lramento	967	835	974	961	1.032	912	926	698	905	1.104	1.601	10.915
Atitude suspeita	273	384	614	699	1.148	991	1.120	978	1096	1.014	1.355	9.672
Pichação	685	643	713	1.148	2.132	1.423	1.109	521	226	242	307	9.149
Disparo de alarme (viola- ção)	2.375	1.895	1.570	391	546	665	446	355	327	361	157	9.088
Vandalismo	791	770	790	874	1.050	1.045	921	703	659	634	648	8.885
Cão solto em via públic a	2.060	1.845	1.679	1.266	1.105	228	388	160	14	8	16	8.769
Transporte coletivo	189	146	284	356	552	1.252	1.505	1.315	973	680	508	7.760
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

FONTE: A Autora (2020)

A oitava análise realizada é no atributo Ocorrencia\_mes à qual foi relacionada com o atributo Atendimento\_ano, e apresenta a relação dos meses que mais aconteceram ocorrências distribuídas pelos anos conforme está apresentado no Gráfico 4.

GRÁFICO 4 - DISTRIBUIÇÃO DAS 10 PRINCIPAIS OCORRENCIA\_MES



FONTE: A Autora (2020)

Percebe-se através do Gráfico 4 uma distribuição uniforme das ocorrências nos meses em relação aos anos, mantendo-se uma média de (21.791) na quantidade de ocorrências registradas.

A nona análise realizada neste capítulo é no atributo Ocorrencia\_dia\_semana, na qual aplicou-se a análise da tabela de frequência absoluta com cruzamento do atributo Atendimento\_Ano, conforme apresentado na Tabela 19, que está ordenada de forma decrescente com a quantidade de ocorrências.

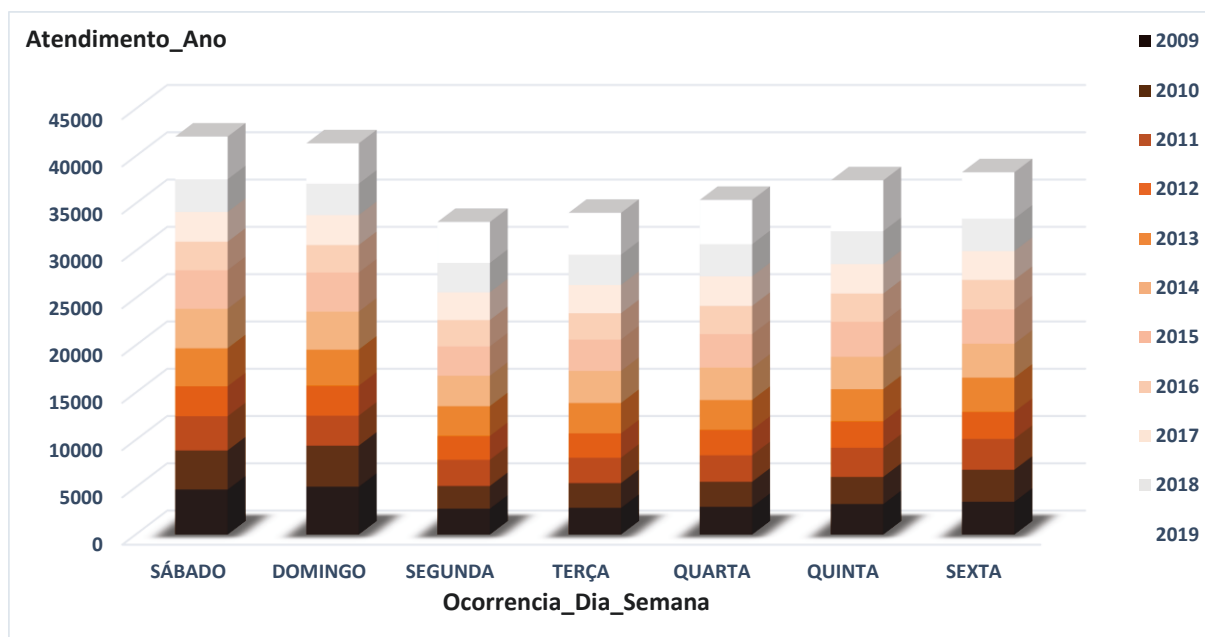
TABELA 19 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE OCORRENCIA\_DIA\_SEMANA POR ANOS

<b>Dia da Semana</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>Total Geral</b>
SÁBADO	4.731	4.145	3.618	3.160	4.022	4.176	4.078	3.013	3.148	3.457	4.508	42.056
DOMINGO	5.031	4.348	3.176	3.157	3.819	4.022	4.131	2.910	3.167	3.322	4.259	41.342
SEGUNDA	2.715	2.404	2.768	2.515	3.151	3.220	3.102	2.788	2.932	3.110	4.327	33.032
TERÇA	2.797	2.628	2.682	2.568	3.213	3.393	3.321	2.781	2.994	3.196	4.420	33.993
QUARTA	2.904	2.661	2.801	2.681	3.161	3.424	3.541	2.983	3.126	3.402	4.677	35.361
QUINTA	3.197	2.879	3.082	2.787	3.407	3.421	3.701	2.991	3.114	3.483	5.367	37.429
SEXTA	3.447	3.389	3.261	2.845	3.630	3.594	3.620	3.121	3.053	3.431	4.893	38.284
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

FONTE: A Autora (2020)

Através da Tabela 19 percebe-se que o dia da semana que mais são registradas ocorrências é o sábado e domingo, ultrapassando (40.000) ocorrências em ambos somado todos os anos, com maior frequência no ano de 2009 porém se manteve uniforme o total de ocorrências nos demais anos, já o dia da semana que menos apresentou ocorrências foi a segunda feira com (33.032) registros somados. Para melhor visualização o Gráfico 5 ilustra a Tabela 19.

GRÁFICO 5 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR DIA DA SEMANA NOS ANOS



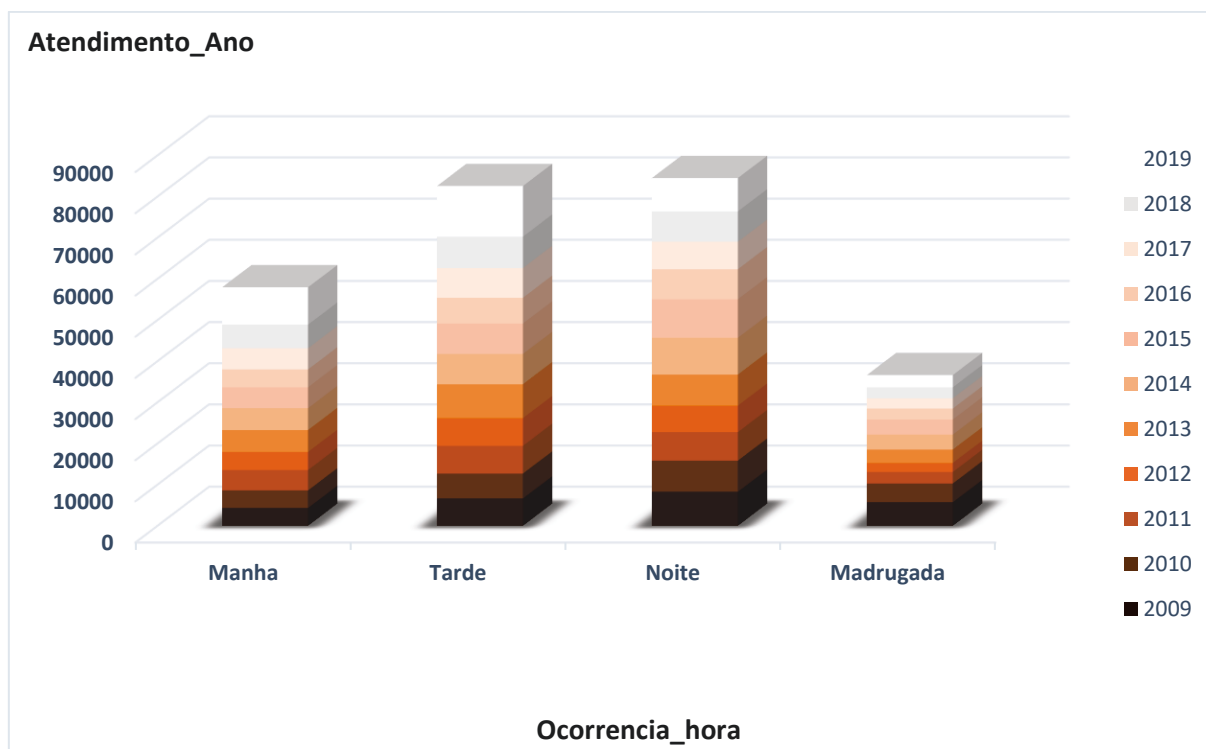
FONTE: A Autora (2020)

Através do Gráfico 5 percebe-se que sábado e domingo concentram a maior parte das ocorrências, porém de segunda a sexta existe uma linha crescente de ocorrências.

A decima análise realizada no capítulo de estatística relaciona o atributo Ocorrencia\_Hora com Atendimento\_Ano, a qual resultou no Gráfico 6.



GRÁFICO 6 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR HORA NOS ANOS



FONTE: A Autora (2020)

Através do Gráfico 6 nota-se que o período em que mais apresentou ocorrências foi a noite, em sequência da tarde, manhã e madrugada. Para melhor análise a Tabela 20 contém de forma detalhada explicando o gráfico.

TABELA 20 - FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE OCORRÊNCIA HORA POR ANOS

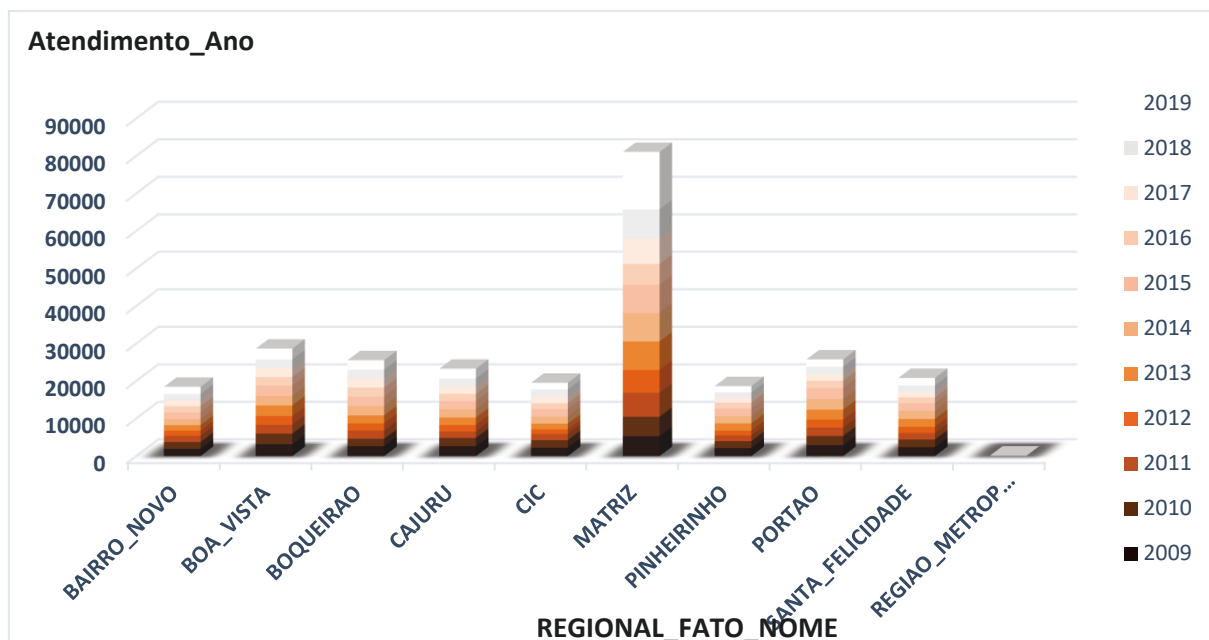
Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
Manhã	4.295	4.278	4.941	4.387	5.327	5.317	5.080	4.336	5.141	5.741	9.078	57.921
Tarde	6.603	6.053	6.728	6.721	8.256	7.352	7.393	6.268	7.229	7.612	1.2306	82.521
Noite	8.230	7.568	6.926	6.418	7.553	8.941	9.366	7.310	6.671	7.381	8.094	84.458
Madrugada	5.694	4.555	2.793	2.187	3.267	3.640	3.655	2.673	2.493	2.667	2.973	36.597
<b>Total Geral</b>	<b>24.822</b>	<b>22454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

FONTE: A Autora (2020)

A grande concentração de ocorrências distribuídas pelos anos está em 2018 e 2019 segundo a Tabela 20, os demais anos mantiveram uma sequência de registros de ocorrências.

A décima primeira análise realizada na base de dados é a relação do atributo Regional\_fato\_nome com Atendimento\_Ano através do Gráfico 7, (9) regionais que pertencem a Curitiba e uma a Região Metropolitana a qual atende em regiões divisas com Curitiba.

GRÁFICO 7 - GRÁFICO DE DISTRIBUIÇÃO DE OCORRÊNCIAS POR REGIONAL NOS ANOS



FONTE: A Autora (2020)

Percebe-se que a regional matriz tem disparadamente mais ocorrências que as demais regionais, para uma melhor análise na Tabela 21 contém de forma detalhada cada regional relacionado com o ano em uma tabela de frequências relativas absoluta.

TABELA 21 - TABELA DE FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DE REGIONAL POR ANOS

Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
BAIRRO_NOVO	1.905	1.796	1.655	1.353	1.519	1.640	1.702	1.698	1.614	1.667	1.856	18.405
BOA_VISTA	3.112	2.840	2.375	2.281	2.867	2.549	2.701	2.274	2.381	2.374	2.855	28.609
BOQUEIRAO	2.575	1.989	2.196	1.887	2.169	2.486	2.486	2.415	2.444	2.365	2.504	25.516
CAJURU	2.571	2.230	1.691	1.739	2.009	2.202	2.141	1.982	1.958	2.139	2.576	23.238
CIC	2.178	2.004	1.637	1.246	1.534	1.892	2.009	1.539	1.715	1.958	1.772	19.484
MATRIZ	5.235	5.206	6.371	6.087	7.630	7.586	7.493	5.574	6.815	7.757	15.284	81.038
PINHEIRINHO	2.091	1.838	1.500	1.338	1.906	1.928	2.029	1.560	1.291	1.467	1.666	18.614
PORTAO	2.853	2.468	2.232	2.114	2.686	2.801	2.917	1.972	1.788	1.966	1.931	25.728
SANTA_FELICIDADE	2.302	2.083	1.731	1.668	2.083	2.166	2.007	1.565	1.507	1.686	1.990	20.788

REGIAO_METROPOLITANA							9	8	21	22	17	77
<b>Total Geral</b>	24.822	22.454	21.388	19.713	24.403	25.250	25.494	20.587	21.534	23.401	32.451	261.497

FONTE: A Autora (2020)

Através dos resultados apresentados na Tabela 21 e no Gráfico 7, percebe-se que a regional Matriz é a que mais apresentou ocorrências registradas pelo fato de ser a região central e a que mais contém bairros disparando com (81.038) registros no total entre os anos, sendo em 2019 com (15.284) ocorrências o ano que mais se teve registros e o ano que menos teve ocorrências foi em 2010 com (5.206) registros. A segunda regional com mais ocorrências é Boa Vista com (28.609) ocorrências somadas com todos os anos, as demais regionais estão com uma média de 20 mil ocorrências. A regional que menos registrou ocorrências é o Bairro novo com (18.405) ocorrências ao total. A região metropolitana contém apenas (77) registros.

Através da análise do atributo Regional\_Fato\_Nome e Atendimento\_Ano na decima primeira análise foi identificado que a regional que mais continha ocorrências era a Matriz, para esta décima segunda análise foram relacionados os atributos Regional\_Fato\_Nome e Atendimento\_Bairro\_Nome filtrados pela regional que mais contém ocorrências conforme apresentado no Tabela 22.

TABELA 22 - FREQUÊNCIA ABSOLUTAS DA DISTRIBUIÇÃO DA REGIONAL MATRIZ POR BAIRROS

<b>MATRIZ</b>	<b>Total</b>
AHÚ	589
ALTO DA GLÓRIA	1.674
ALTO DA RUA XV	1.790
BATEL	1.980
BIGORRILHO	2.764
BOM RETIRO	525
CABRAL	2.026
CENTRO	44.160
CENTRO CÍVICO	2.567
CRISTO REI	998
HUGO LANGE	312
JARDIM BOTÂNICO	4.247
JARDIM SOCIAL	372
JUVEVÊ	593
MERCÊS	1.955

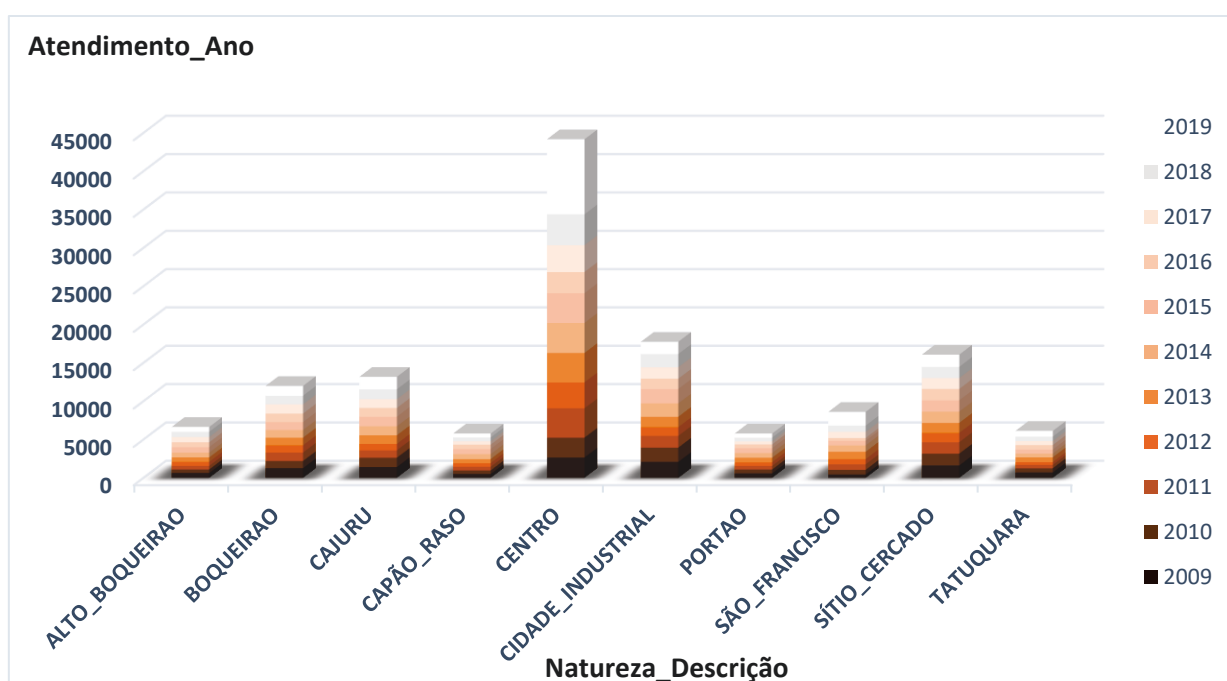
PRADO VELHO	1.569
REBOUÇAS	4.325
SÃO FRANCISCO	8.592
<b>Total Geral</b>	<b>81.038</b>

FONTE: A Autora (2020)

A regional Matriz contém (18) bairros conforme apresentados na Tabela 22, percebe-se que o bairro que mais contém ocorrências é o centro com um total de (44.160) ocorrências já o bairro que menos se apresentou ocorrências foi no Hugo Lange com (312) ocorrências.

A décima terceira análise realizada foi no atributo Atendimento\_bairro\_nome, o atributo contém 133 bairros que estão distribuídos entre Curitiba e alguns da região metropolitana, para esta análise foram selecionados os 10 bairros que mais tiveram ocorrências apresentados no Gráfico 8.

GRÁFICO 8 - DISTRIBUIÇÃO DE OCORRÊNCIAS POR BAIRROS NOS ANOS



FONTE: A Autora (2020)

Através do Gráfico 8 o bairro que mais teve ocorrências foi o centro com (44.160) registros somados entre todos os anos com maior incidência em 2019, já o bairro que menos teve ocorrências no ranking de 10 bairros com mais ocorrências foi

o Capão Raso com (5768) ocorrências somadas no total. Para melhor análise, a Tabela 23 contém informações complementares do gráfico ordenados de forma decrescente, para uma melhor análise a tabela completa está no apêndice P.

TABELA 23 - TABELA FREQUÊNCIAS ABSOLUTAS DA DISTRIBUIÇÃO DE BAIRROS POR ANO

Bairros	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
CENTRO	2.633	2.595	3.833	3.373	3.839	3.921	3.874	2.768	3.507	4.024	9.793	44.160
CIDADE INDUSTRIAL	2.068	1.859	1.527	1.133	1.368	1.746	1.865	1.358	1.497	1.724	1.590	17.735
SÍTIO CERCADO	1.600	1.546	1.488	1.202	1.322	1.468	1.469	1.487	1.430	1.452	1.597	16.061
CAJURU	1.396	1.213	962	838	1.125	1.176	1.239	1.153	1.151	1.288	1.624	13.165
BOQUEIRÃO	1.246	939	1.096	936	1.005	1.008	1.027	1.115	1.211	1.114	1.280	11.977
SÃO FRANCISCO	421	618	715	668	968	825	604	364	805	821	1.783	8.592
ALTO BOQUEIRÃO	620	465	465	509	594	624	706	643	687	688	628	6.629
TATUQUARA	683	557	430	358	632	490	525	583	527	584	742	6.111
PORTÃO	563	516	464	465	600	616	635	487	403	485	535	5.769
CAPÃO RASO	481	472	486	459	531	648	667	571	437	498	518	5.768
Total Geral	24.822	22.454	21.388	19.713	24.403	25.250	25.494	20.587	21.534	23.401	32.451	261.497

FONTE: A Autora (2020)

Através da Tabela 23 complementando a análise do Gráfico 8 observa-se que o bairro “Centro” que mais teve ocorrências concentrou a maior parte das ocorrências no ano de 2019 com (9793) registros. E o bairro que menos teve ocorrências neste ranking de 10 Bairros com mais ocorrências está o bairro Capão Raso com (5768) ocorrências, concentrando a maior parte dos registros em 2015 com (667) registros.

A décima quarta e última análise realizada no capítulo foi a relação do atributo Natureza1\_Descricao e Subcategoria1\_Descricao, para esta análise utilizou-se o ranking das 10 descrições que mais tiveram ocorrências na base de dados sendo a Abordagem; Alarmes; Animais; Apoio; Atitude suspeita; Dano; Invasão; Perturbação do sossego; Roubo e Substância Ilícita. Através da relação entre os dois atributos resultou em (78) subcategorias utilizadas em (10) descrições, para melhor visualização foram selecionadas as 20 primeiras subcategorias conforme a Tabela 24 ordenados de forma decrescente, a tabela completa com todas subcategoria está presente no apêndice Q.

TABELA 24 - FREQUÊNCIA ABSOLUTAS DAS 10 PRINCIPAIS DESCRIÇÕES POR SUBCATEGORIA

Subcategoria	Abordagem	Alarmes	Animais	Apoio	Atitude suspeita	Dano	Invasão	Perturbação do sossego	Roubo	Substância Ilícita	Total Geral
Abordagem	19.216										19.216
Perturbação do sossego								18.990			18.990
Uso de substância ilícita										11.684	11.684
Apoio ao cidadão PRESTAÇÃO DE SOCORRO/SALVAMENTO				10.915							10.915
Atitude suspeita					9.672						9.672
Pichação						9.149					9.149
Disparo de Alarme (violação)		9.088									9.088
Vandalismo						8.885					8.885
Cão solto em via pública			8.769								8.769
Transporte Coletivo									7.452		7.452
Desordem						6.136					6.136
Transeunte									6.008		6.008
Arrombamento						5.830					5.830
Apoio ao SAMU				5.629							5.629
Invasão de equipamento/patrimônio público							5.620				5.620
Apoio à FAS				4.107							4.107
Invasão ao transporte coletivo							4.100				4.100
Apoio ao cidadão - ORIENTAÇÃO				3.373							3.373
Maus tratos a animais			3.073								3.073
Tráfico de drogas										2.960	2.960
<b>Total Geral</b>	<b>19.216</b>	<b>9.905</b>	<b>12.993</b>	<b>39.733</b>	<b>9.672</b>	<b>33.005</b>	<b>11.271</b>	<b>18.990</b>	<b>15.788</b>	<b>16.346</b>	<b>187.245</b>

FONTE: A Autora (2020)

Através do Tabela 24, percebe-se a relação de algumas subcategorias com as descrições que trazem destaque por exemplo a descrição animais no ranking das 20 subcategorias mais utilizadas contém mais ocorrências em “Cão solto em via pública” e “Maus tratos a animais”, outra descrição em destaque é o “Dano” nesta descrição as três subcategorias com mais registros é “Pichação, Vandalismo, Desordem e Arrombamento”.

Finalizando o capítulo de estatística descritiva, a Tabela 25 contém de forma resumida as análises realizadas no percorrer do capítulo, sendo analisadas as variações dos atributos, máximo e mínimo.

TABELA 25 - DISTRIBUIÇÃO DE VARIAÇÃO, MÍNIMO E MÁXIMO DE OCORRÊNCIAS

Atributo	Variação	Máximo de Ocorrências	Mínimo de Ocorrências
Atendimento_ano	De 2009 até 2019	2019 (32.451)	2012 (19713)
Origem_chamado_descricao	25 origens	153 - Guarda Municipal (101.286)	Central de flagrantes (2)
Secretaria sigla	38 Secretarias	SMDT (13.7718)	SEPE e SERCOM (2)
Flag_flagrante	Sim ou não	NÃO (24.8714)	SIM (12783)
Natureza1_defesa_civil	Sim ou não	NÃO (25.3109)	SIM (8388)
Natureza1_descricao	167 Descrições	Apoio (39.733)	Corrupção ativa, Liberação de pessoa presa, Óbito, Prostituição, Quadrilha ou bando, Rompimento de Barragem e Rufianismo (1)
Subcategoria1_descricao	283 Subcategorias	Abordagem (19.216)	Corpo_de_Bombeiros, Corrupção_ativa, Defesa_Civil, Falsificação_de_documento_Publico, Gazuas/chaves_falsas, Liberação_de_pessoa_presa/apreendida_por_recusa_no_recebimento_pela_DP, Óbito_(Defesa_Civil), Polícia_Civil, Prostituição, Quadrilha_ou_bando, Rompimento_de_Barragem, Rufianismo e Vender_bebida_alcoólica_a_menor_de_18_anos
Ocorrencia_mes	12 meses	Maio (22.892)	Fevereiro (20007)
Ocorrencia_dia_semana	De Segunda até Sábado	Sábado (42.056)	Segunda (33032)
Ocorrencia_hora	Períodos (Manhã, Tarde, Noite, Madrugada)	Noite (84.458)	Madrugada (36597)
Regional_fato_nome	9 Regionais em Curitiba e Região Metropolitana	Matriz (81.038)	Bairro Novo (18405) e Região metropolitana (77)
Atendimento_bairro_nome	136 Bairros considerando a Região Metropolitana	Centro (44.160)	Caximba – Curitiba (213) Vila Perneta – Região metropolitana (1)

FONTE: A Autora (2020)

### 4.3 MINERAÇÃO DE DADOS

Após a realização do pré-processamento com a limpeza dos dados, tratamento da base de dados e estatística descritiva o presente capítulo de mineração de dados tem o objetivo de descobrir padrões na base de dados da Guarda Municipal de Curitiba com a aplicação do algoritmo Árvore de Decisão e o Naïve Bayes.

Conforme apresentado no capítulo de metodologia, a ferramenta utilizada para o presente trabalho na aplicação dos algoritmos foi o Rstudio, porém para aplicação do algoritmo foram realizados dois processamentos em cada algoritmo, sendo um com os dados da Guarda Municipal de Curitiba agrupados às descrições de forma manual e o outro arquivo de dados agrupados pelo K-Means.

Antes de ser carregado a base de dados, foi definido o atributo meta e os demais atributos que compõem a aplicação dos algoritmos. O atributo meta utilizado foi a natureza1\_descricao com as instâncias agrupadas sendo renomeadas para grupodescricao. Foi necessário realizar o agrupamento pois o atributo natureza descrição continha 167 instâncias únicas e o processamento do algoritmo em bases deste tamanho leva muito tempo para ser processado talvez até dias. Portanto, com a definição deste atributo meta o objetivo é identificar se existe relação e um padrão entre os atributos selecionados da base. Os atributos selecionados para compor os algoritmos foram: regional\_fato\_nome, atendimento\_ano, ocorrencia\_dia\_semana, ocorrencia\_hora, ocorrencia\_mes, flag\_flagrante, naturezasim\_defesa\_civil, origem\_chamado\_descricao com o atributo meta codigodescricao.

FIGURA 18 - SCRIPT DE CARGA DE DADOS

```
##### Carregando os dados #####
dadosAgrManual <- read.csv2('Guarda_manual.csv')
dadosAgrKmeans <- read.csv2('Guarda_Kmeans.csv')
```

FONTE: A Autora (2020)

Foi realizada a carga de dados das duas bases de acordo com a Figura 18, a primeira carga realizada é dos dados agrupados de forma manual, nomeadas como "dadosAgrManual" e a segunda carga com os dados agrupados pelo k-Means com o nome "dadosAgrKmeans".



Para a realização da aplicação dos algoritmos, o capítulo de mineração foi dividido em 2 tópicos sendo um sobre a Árvore de Decisão e o outro sobre Naïve Bayes.

#### 4.3.1 Árvore de Decisão

Após o arquivo de dados ser carregado de acordo com a Figura 18 foi instalado os pacotes necessários para realizar a Árvore através do método Part conforme apresentado na Figura 19.

FIGURA 19 - SCRIPT DE INSTALAÇÃO E EXECUÇÃO DE PACOTES DA ÁRVORE DE DECISÃO

```
#####Instalando pacotes #####
install.packages("data.table")
install.packages("rpart")
install.packages("rpart.plot")
install.packages("prp")

##### Carregando pacotes #####
library(data.table)
library(rpart)
library(rpart.plot)
library(prp)

#Esta função data.table possibilita mais funções que a função data.frame (padrão do R)
Manual <- data.table(dadosAgrManual)#Dados com o Agrupamento Manual
K_means <- data.table(dadosAgrKmeans)#Dados com o agrupamento do K-Means
```

FONTE: A Autora (2020)

Foram instalados e executados os pacotes (data.table; rpart; rpart.plot e prp), e após a execução foi aplicado a função data.table em ambas as bases de dados. Essa função disponibiliza mais atribuições do que o R disponibiliza no básico, renomeado os arquivos para (Manual) e (K\_Means) de acordo com o recorte do script na Figura 19.

Para execução do script do algoritmo foram considerados os atributos descritos no capítulo de mineração sendo: regional\_fato\_nome, atendimento\_ano, ocorrencia\_dia\_semana, ocorrencia\_hora, ocorrencia\_mes, flag\_flagrante, naturezasim\_defesa\_civil, origem\_chamado\_descricao com o atributo meta já descrito codigodescricao. O script executado do algoritmo está presente na Figura 20, porém como dito antes, foi aplicado tanto para a base Manual quanto Means. Portanto de acordo com a figura foi executado tanto para o arquivo de dados manual quanto para o k-Means.

FIGURA 20 - SCRIPT DO ALGORITMO DE ÁRVORE DE DECISÃO

```
#Algoritmo da Arvore de Decisão
Arvore <- rpart(CODIGODESCRICAO ~ REGIONAL_FATO_NOME + ATENDIMENTO_ANO + OCORRENCIA_DIA_SEMANA +
OCORRENCIA_HORA + OCORRENCIA_MES + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL + ORIGEM_CHAMADO_DESCRICAO ,
data = Manual, method = 'class')
```

FONTE: A Autora (2020)

Após executado o script da Figura 20 ambos os arquivos levaram menos de 1 minuto de processamento, o próximo passo foi a visualização para a análise do resultado, a Figura 21 contém os scripts necessários para visualização da árvore.

FIGURA 21 - SCRIPT DE VISUALIZAÇÃO DA ÁRVORE

```
#Visualização da arvore
Arvore #visualiza os resultados no console
prp(Arvore,type=4, extra=102) #visualiza a arvore com a probabilidade e rótulos de divisão abaixo dos rótulos dos nós
```

FONTE: A Autora (2020)

Através dos scripts apresentados na Figura 21 obtiveram-se 4 visualizações, sendo 2 da base de dados Manual e 2 da base de dados K-Means. A primeira visualização presente no código “árvore” apresenta o resultado no console, e nas duas bases apresentaram bastante diferença pelo simples fato que ambas têm agrupamentos diferentes. Na Figura 22 é apresentado o resultado do console da base de dados agrupada manualmente:

FIGURA 22 - RESULTADO DA ÁRVORE NO CONSOLE DA BASE DE DADOS MANUAL

```

n= 26149/

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 261497 193144 GRUPO_3 (0.087 0.23 0.26 0.25 0.17)
  2) ORIGEM_CHAMADO_DESCRICAO=A_VIATURA,AOS_GMS,CASA_DA_MULHER_BRASILEIRA,CCO_BARIGUI,CCO_MUNICIPAL_URBS,CCO_PUC,CCO_RUAS_DA_CIDADANIA,CCO_SÍTIO_CERCADO,CCO_URBS,CENTRAL_DE_FLAGRANTES,EMAIL,FONE,IMEDIATA,NRGM,OFÍCIO,SAMU,SIGA,TOTEM_RUA_SÃO_FRANSICO 117607 77664 GRUPO_2 (0.027 0.34 0.32 0.09 0.22)
    4) ORIGEM_CHAMADO_DESCRICAO=A_VIATURA,AOS_GMS,CASA_DA_MULHER_BRASILEIRA,CCO_BARIGUI,CCO_MUNICIPAL_URBS,CCO_PUC,CCO_RUAS_DA_CIDADANIA,CCO_SÍTIO_CERCADO,CCO_URBS,CENTRAL_DE_FLAGRANTES,EMAIL,FONE,IMEDIATA,NRGM,OFÍCIO,SAMU,TOTEM_RUA_SÃO_FRANSICO 111683 71776 GRUPO_2 (0.029 0.36 0.34 0.094 0.18)
      8) ORIGEM_CHAMADO_DESCRICAO=CCO_MUNICIPAL_URBS,EMAIL,SAMU 2878 114 GRUPO_2 (0.0049 0.96 0.0083 0.0083 0.018) *
      9) ORIGEM_CHAMADO_DESCRICAO=A_VIATURA,AOS_GMS,CASA_DA_MULHER_BRASILEIRA,CCO_BARIGUI,CCO_PUC,CCO_RUAS_DA_CIDADANIA,CCO_SÍTIO_CERCADO,CCO_URBS,CENTRAL_DE_FLAGRANTES,FONE,IMEDIATA,NRGM,OFÍCIO,TOTEM_RUA_SÃO_FRANSICO 108805 71288 GRUPO_3 (0.029 0.34 0.34 0.096 0.19)
        18) NATUREZASIM_DEFESA_CIVIL=SIM 1927 588 GRUPO_1 (0.69 0.2 0 0.049 0.052) *
        19) NATUREZASIM_DEFESA_CIVIL=NAO 106878 69361 GRUPO_3 (0.017 0.34 0.35 0.097 0.19)
          38) ATENDIMENTO_ANO=F,G,H,I,J,K 65759 38806 GRUPO_3 (0.012 0.33 0.41 0.083 0.16) *
          39) ATENDIMENTO_ANO=A,B,C,D,E 41119 26338 GRUPO_2 (0.026 0.36 0.26 0.12 0.24) *
        5) ORIGEM_CHAMADO_DESCRICAO=SIGA 5924 96 GRUPO_5 (0 0.0061 0.00068 0.0095 0.98) *
      3) ORIGEM_CHAMADO_DESCRICAO=CENTRAL_DE_ATENDIMENTO_PREFEITURA,CIOSP,DEFESA_CIVIL,G5,GUARDA_MUNICIPAL,REDE_SOCIAL,VEPPER 143890 89381 GRUPO_4 (0.14 0.14 0.21 0.38 0.13)
        6) ORIGEM_CHAMADO_DESCRICAO=CENTRAL_DE_ATENDIMENTO_PREFEITURA,DEFESA_CIVIL 14191 3852 GRUPO_1 (0.73 0.1 0.009 0.14 0.022) *
        7) ORIGEM_CHAMADO_DESCRICAO=CIOSP,G5,GUARDA_MUNICIPAL,REDE_SOCIAL,VEPPER 129699 77183 GRUPO_4 (0.071 0.15 0.24 0.4 0.14)
          14) OCORRENCIA_HORA=Manha,Noite,Tarde 104070 68821 GRUPO_4 (0.084 0.16 0.26 0.34 0.15)
            28) NATUREZASIM_DEFESA_CIVIL=SIM 3476 1055 GRUPO_1 (0.7 0.2 0 0.03 0.077) *
            29) NATUREZASIM_DEFESA_CIVIL=NAO 100594 65450 GRUPO_4 (0.063 0.16 0.27 0.35 0.16)
              58) ATENDIMENTO_ANO=F,G,H,I,J,K 59787 39151 GRUPO_3 (0.042 0.17 0.35 0.27 0.17) *
              59) ATENDIMENTO_ANO=A,B,C,D,E 40807 21929 GRUPO_4 (0.094 0.14 0.16 0.46 0.14) *
          15) OCORRENCIA_HORA=Madrugada 25629 8362 GRUPO_4 (0.02 0.088 0.13 0.67 0.085) *

```

FONTE: A Autora (2020)

A visualização do resultado pelo console da base de dados manual detalha a relação dos atributos e as ramificações de cada grupo, a sua interpretação fica melhor visualizada através de uma árvore.

A Figura 23 apresenta o resultado do console da base de dados agrupada pelo k-means:

FIGURA 23 - RESULTADO DA ÁRVORE NO CONSOLE DA BASE DE DADOS K-MEANS

```

n= 261497

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 261497 118419 G4 (0.015 0.044 0.13 0.55 0.27)
  2) ORIGEM_CHAMADO_DESCRICAO=A_VIATURA,AOS_GMS,CASA_DA_MULHER_BRASILEIRA,CCO_BARIGUI,CCO_MUNICIPAL_URBS,CCO_PUC,CCO_RUAS_DA_CIDADANIA,CCO_SÍTIO_CERCADO,CCO_URBS,CENTRAL_DE_FLAGRANTES,CIOSP,DEFESA_CIVIL,EMAIL,FONE,GG5,GUARDA_MUNICIPAL,IMEDIATA,NRGM,OFÍCIO,REDE_SOCIAL,SAMU,TOTEM_RUA_SÃO_FRANSICO,VEPPER 242094 100289 G4 (0.016 0.047 0.13 0.59 0.22)
    4) NATUREZASIM_DEFESA_CIVIL=SIM 6262 3611 G3 (0.16 0.41 0.42 0 0) *
    5) NATUREZASIM_DEFESA_CIVIL=NAO 235832 94027 G4 (0.012 0.037 0.12 0.6 0.23) *
  3) ORIGEM_CHAMADO_DESCRICAO=CENTRAL_DE_ATENDIMENTO_PREFEITURA,SIGA 19403 3577 G5 (0.0023 0.007 0.11 0.066 0.82)
    6) NATUREZASIM_DEFESA_CIVIL=SIM 2126 127 G3 (0.012 0.048 0.94 0 0) *
    7) NATUREZASIM_DEFESA_CIVIL=NAO 17277 1451 G5 (0.0012 0.002 0.0072 0.074 0.92) *

```

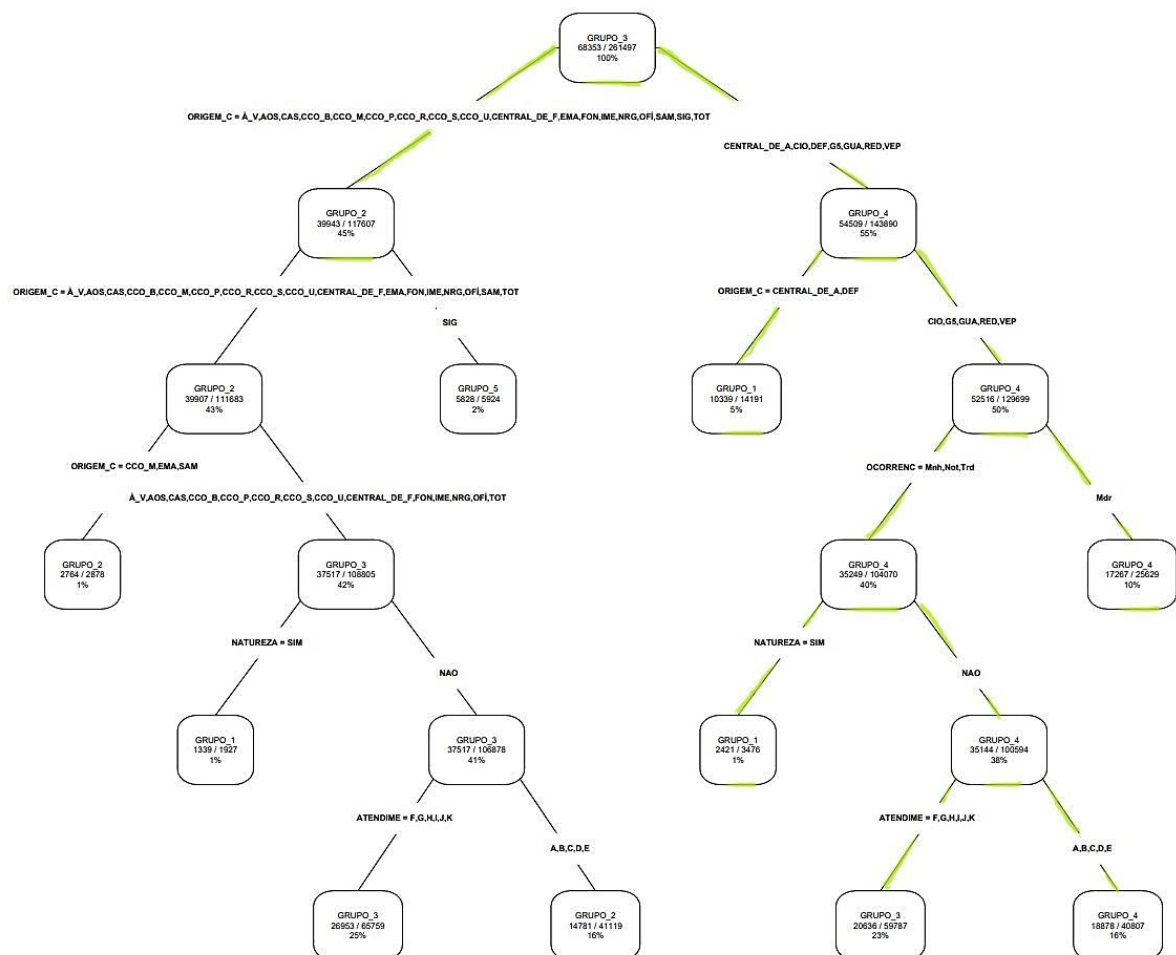
FONTE: A Autora (2020)

A visualização do resultado pelo console da base de dados pelo k-Means obteve menos ramificações que a base de dados manual, assim como já mencionado na análise do console manual para uma melhor interpretação fica melhor visualizada através de uma árvore.

A segunda visualização colocada nos scripts da Figura 21 é o prp (Árvore, type=4, extra=102) responsável por gerar a árvore. Através dessa função obtiveram-se como resultado e interpretação as definições prp com o Type configurado igual a 4 rotula os nós além das folhas, o extra igual a 102 apresenta as probabilidades através do código 10 e o 2 apresenta a taxa de classificação de cada nó.

A Figura 24 apresenta a árvore de decisão dos dados agrupados Manual:

FIGURA 24 - ÁRVORE DE DECISÃO COM DADOS AGRUPADOS MANUAL



FONTE: A Autora (2020)

A análise da árvore de decisão na Figura 24 foi selecionado o lado direito para uma análise mais profunda devido ao tamanho da árvore, portanto, para estudo deste capítulo foi analisada a primeira ramificação uma vez que esta concentra o maior número de ocorrências originada do nó raiz e toda sendo toda a ramificação do lado direito conforme destacado na figura em cor verde.

O nó raiz se inicia no Grupo 3 a qual concentra a maior partes das ocorrências atendidas da base, através da árvore no nó raiz ele se divide em duas ramificações sendo a primeira do lado esquerdo com a distribuição do atributo origem (*à\_viatura,aos\_gms,casa\_da\_mulher\_brasileira,cco\_barigui,cco\_municipal\_urbs,cco\_puc,cco\_ruas\_da\_cidadania,cco\_sítio\_cercado,cco\_urbs,central\_de\_flagrantes,email,fone,imediata,nrgm,of\_ício,samu,siga,totem\_rua\_são\_fransico*) e o nó referente a este galho foi o grupo\_2 com 45%. Já a ramificação do lado direito com o atributo origem (*central\_de\_atendimento\_prefeitura, ciosp, defesa\_civil, g5, guarda\_municipal, rede\_social, vepper*) ficou classificado com o nó do grupo\_4 com 55%.

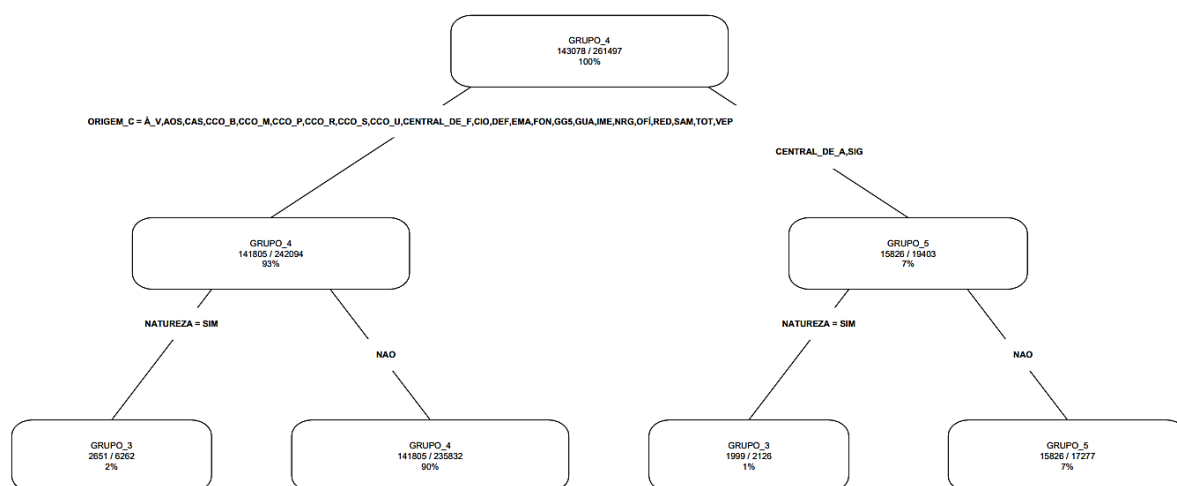
A partir do nó do grupo \_4 surgem duas ramificações, a primeira do lado direito para a primeira folha e a segunda ramificação tem continuação. A ramificação foi dividida através do atributo origem. A folha do lado esquerdo concentrando as origens (*central\_de\_atendimento\_prefeitura* e *defesa\_civil*) mais presente no grupo\_1 com 5% das ocorrências. A segunda ramificação conectada ao nó raiz do lado direito contém as origens (*ciosp, g5, guarda\_municipal, rede\_social, vepper*) e está presente no nó grupo\_4 com 50% de ocorrências. Este segundo nó da ramificação do lado direito foi classificado através do atributo *Ocorrencia\_Hora*, a ramificação do lado direito concentra os períodos (manhã, tarde e Noite) concentrando o grupo\_4 com 40% e a segunda ramificação concentra o período da madrugada trazendo a folha final com o grupo\_4 a 10%.

O terceiro nó conectado ao nó raiz da ramificação do lado direito se divide em duas ramificações classificadas através do atributo *Natureza\_defesa\_civil*, a ramificação do lado esquerdo com a variável (SIM) do atributo já leva direto para a folha final apresentando como resultado o grupo\_1 e 1%, já a ramificação do lado direito com a variável(não) se conecta ao nó do grupo\_4 igual a 38% e ele está dividido em duas ramificações que dão direto as folhas, estas ramificações estão classificadas com o atributo *ATENDIMENTO\_ANO*, a ramificação do lado esquerdo se concentra os anos (F,G,H,I,J,K) sendo os períodos mais atuais e dá a folha grupo\_3 com 23%,

já a ramificação do lado direito se concentra os anos (A,B,C,D,E) que são os períodos mais antigos e dá na folha cujo grupo é 4 com 16% das ocorrências.

A Figura 25 apresenta a árvore de decisão dos dados agrupados pelo K-Means:

FIGURA 25 - ÁRVORE DE DECISÃO COM DADOS AGRUPADOS K-MEANS



FONTE: A Autora (2020)

Através Figura 25 percebe-se que a Árvore de decisão com os dados agrupados K-Means tiveram bastante diferença comparando com a árvore da Figura 24. Primeiramente pelo tamanho esta contém bem menos nós e folhas. A maior concentração dos dados está presente no grupo\_4 com (143.078) sendo o nó raiz, deste nó surgiu 2 ramificações considerando o atributo origem\_chamado\_descricao a primeira ramificação do lado esquerdo concentra a maior parte das origens sendo (à\_viatura, aos\_gms, casa\_da\_mulher\_brasileira, cco\_barigui, cco\_municipal\_urbs, cco\_puc, cco\_ruas\_da\_cidadania, cco\_sítio\_cercado, cco\_urbs, central\_de\_flagrantes, ciosp, defesa\_civil, email, fone, g5, guarda\_municipal, imediata, nrgm, ofício, rede\_social, samu, totem\_rua\_são\_fransico, vepper) e esta ramificação ficou conectada ao nó do grupo\_4 com 93% das ocorrências. O grupo 4 se divide em duas ramificações que dão direto nas folhas, porém estas ramificações foram classificadas pelo atributo natureza\_defesa\_civil, a ramificação do lado esquerdo com a natureza igual a (SIM) resultou na folha de grupo\_3 igual a 2%, já a ramificação do lado direito com a variável (NÃO) do atributo natureza resultou a folha com o grupo\_4 com 90% das ocorrências.

A segunda ramificação que está conectada ao nó raiz do lado direito considerou a Origem (central\_de\_atendimento\_prefeitura e siga) conectando ao nó do grupo\_5 que representa (15826) dos dados 7%. Deste nó resultaram em mais duas ramificações considerando o atributo natureza descrição, a ramificação do lado esquerdo considerou o atributo natureza “sim” e a folha o grupo\_3 com 1% e a ramificação do lado direito considerou o atributo natureza “não” é a folha resposta o grupo\_5 com 7%.

Após a visualização da árvore foram executados os algoritmos presentes na Figura 26 a qual gera a matriz confusão para ambas as bases de dados.

FIGURA 26 - SCRIPT DE GERAÇÃO DA MATRIZ CONFUSÃO DA ÁRVORE DE DECISÃO

```
#Matriz confusão base de dados Manual
matriz <- predict(Arvore, Manual, type = "class")
table( Manual$CODIGODESCRICAO, matriz)

#Matriz confusão base de dados K_means
matriz <- predict(Arvore, K_means, type = "class")
table( K_means$CODIGODESCRICAO, matriz)
```

FONTE: A Autora (2020)

A partir da função “predict” foi executado o resultado da árvore junto da tabela de dados (Manual e k-Means) com a definição do tipo igual a “Class”, o segundo código executado é o “table” ele relaciona o resultado da função do predict anterior cujo nome está matriz com o atributo meta “codigodescricao”, gerando a matriz confusão. A Figura 27 está apresentando a matriz confusão dos dados agrupados manual.

FIGURA 27 - MATRIZ CONFUSÃO DA ÁRVORE DOS DADOS AGRUPADOS MANUAL

	matriz				
	GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
GRUPO_1	14099	1066	3295	4331	0
GRUPO_2	2492	17545	32225	8168	36
GRUPO_3	128	10588	47589	10044	4
GRUPO_4	2193	4957	21723	36145	56
GRUPO_5	682	9841	20714	7748	5828

FONTE: A Autora (2020)

De acordo com a matriz confusão, logo percebe-se que ela teve os valores dispersos seguindo a análise na diagonal e dentro de cada grupo. A matriz confusão

é analisada na diagonal portanto a concentração dos valores tem que estar maior no meio diferente do que se vê na Figura 27. A taxa de acerto se dá através da soma dos valores que estão em diagonal dividindo pelo total de ocorrências (linhas) portanto através da aplicação da árvore na base de dados com agrupamento manual teve  $(14099 + 17545 + 47589 + 36145 + 5828/261497) = 46\%$  de taxa de acerto.

A Figura 28 está apresentando a matriz confusão dos dados agrupados pelo K-Means.

FIGURA 28 - MATRIZ CONFUSÃO DA ÁRVORE DOS DADOS AGRUPADOS KMEANS

	matriz				
	GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
GRUPO_1	0	0	1049	2838	20
GRUPO_2	0	0	2689	8712	34
GRUPO_3	0	0	4650	28991	124
GRUPO_4	0	0	0	141805	1273
GRUPO_5	0	0	0	53486	15826

FONTE: A Autora (2020)

A matriz confusão dos dados agrupados pelo K-Means da Figura 28 diferentemente da matriz anterior teve grupos zerados na classificação da árvore, porém percebe-se que a maior concentração dos dados está na diagonal principalmente no grupo\_4. A Taxa de acerto foi igual a  $(0 + 0 + 4650 + 141805 + 15826/261497) = 62\%$  maior que a matriz confusão manual (46%).

#### 4.3.2 Naïve Bayes

Após a execução dos scripts de carga dos dados na Figura 18, foi realizado a instalação e execução dos pacotes necessários para o algoritmo Naïve Bayes de acordo com a Figura 29.



FIGURA 29 - SCRIPT DE INSTALAÇÃO E EXECUÇÃO DE PACOTES DO NAÏVE BAYES

```
#####Instalando pacotes #####
install.packages("data.table")
install.packages("dplyr")
install.packages("psych")
install.packages("caret")
install.packages("stringr")
install.packages("naivebayes")
install.packages("e1071")

##### Carregando pacotes #####
library(data.table)
library(dplyr)
library(psych)
library(caret)
library(stringr)
library(naivebayes)
library(e1071)

#Esta função data.table possibilita mais funções que a função data.frame (padrão do R)
Manual <- data.table(dadosAgrManual)#Dados com o Agrupamento Manual
K_means <- data.table(dadosAgrKmeans)#Dados com o agrupamento do K-Means
```

FONTE: A Autora (2020)

Os pacotes instalados foram: (data. table; dplyr; psych; caret; stringr; naivebayes e e1071). Após executados os pacotes, a base de dados “dadosAgrManual” e “dadosAgrKmeans” foi aplicado na função data. table e renomeado os arquivos para (Manual) e (K\_means) de acordo com o recorte do script na Figura 29.

Para execução do algoritmo foram considerados os mesmos atributos na árvore de decisão descritos no capítulo de mineração, sendo regional\_fato\_nome, atendimento\_ano, ocorrencia\_dia\_semana, ocorrencia\_hora, ocorrencia\_mes, flag\_flagrante, naturezasim\_defesa\_civil, origem\_chamado\_descricao com o atributo meta já descrito codigodescricao. A Figura 30 contém o script executado do algoritmo, foi aplicado para a base de dados Manual quanto K\_means, portanto de acordo com a figura de exemplo, o arquivo de dados manual foi executado com o k-Means também.

FIGURA 30 - SCRIPT DO ALGORITMO DE NAÏVE BAYES

```
#Algoritmo da Naive Bayes
Bayes <- naiveBayes(CODIGODESCRICAO ~ ATENDIMENTO_ANO + OCORRENCIA_MES + OCORRENCIA_DIA_SEMANA
+ OCORRENCIA_HORA + REGIONAL_FATO_NOME + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL
+ ORIGEM_CHAMADO_DESCRICAO ,data = Manual)
```

FONTE: A Autora (2020)

Após ser executado o script da Figura 30 nas bases de dados (Manual e k-means) ambos levaram menos de 1 minuto de processamento, para a visualização dos resultados foi executado somente o comando “Bayes” e apresentados os resultados.

O resultado do algoritmo aplicado a base de Dados Manual traz probabilidades, a primeira análise resultante do algoritmo é a distribuição dos grupos presente na figura 31.

FIGURA 31 - DISTRIBUIÇÃO DE PROBABILIDADES POR GRUPO NA BASE MANUAL DO NAÏVE BAYES

GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
0.08715588	0.23123019	0.26139114	0.24885180	0.17137099

FONTE: A Autora (2020)

Destaca-se que os grupos 3, 4 e 2 foram os que obtiveram o resultado maior conforme apresentado na Figura 31.

Através do algoritmo o atributo atendimento\_ano tem a concentração das melhores probabilidades no Grupo\_1 e Grupo\_3 conforme a Figura 32.

FIGURA 32 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ATENDIMENTO\_ANO NA BASE MANUAL

ATENDIMENTO_ANO										
	A	B	C	D	E	F	G	H	I	J
GRUPO_1	0.15207757	0.12768198	0.13733491	0.10517309	0.13610636	0.10350577	0.07823264	0.04168312	0.03045062	0.04102497
GRUPO_2	0.07652234	0.07139549	0.07352893	0.06795555	0.07834155	0.08512222	0.08229418	0.08021037	0.09182020	0.11915787
GRUPO_3	0.03508259	0.04276330	0.05275555	0.05875382	0.07401285	0.10126842	0.12365222	0.11011953	0.11229939	0.10930025
GRUPO_4	0.14801610	0.12684021	0.07715831	0.08470357	0.10850724	0.10474229	0.10030120	0.06617082	0.06498755	0.06034668
GRUPO_5	0.10485797	0.09037556	0.11570303	0.08209671	0.09916765	0.08939370	0.08381496	0.06591837	0.07549149	0.08620266

FONTE: A Autora (2020)

Observa-se na Figura 32 que os anos de 2009 até 2013 (A até E) concentram-se no grupo\_1 e os anos de 2015 até 2017 e 2019 (G, H, I, K) concentram-se no Grupo\_3.

Através da Figura 33 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Ocorrência\_Mes.

FIGURA 33 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRÊNCIA\_MES NA BASE MANUAL

OCORRENCIA_MES												
	JANEIRO	FEVEREIRO	MARÇO	ABRIL	MAIO	JUNHO	JULHO	AGOSTO	SETEMBRO	OUTUBRO	NOVEMBRO	DEZEMBRO
GRUPO_1	0.10420780	0.09301917	0.08297135	0.08213769	0.08595498	0.07147558	0.07634593	0.07371331	0.07634593	0.09064982	0.07687245	0.08630600
GRUPO_2	0.07304932	0.07137896	0.08083882	0.08356762	0.09013330	0.07944961	0.08315417	0.09226673	0.08833063	0.08715642	0.08277379	0.08790064
GRUPO_3	0.07300338	0.07309116	0.08416602	0.08254210	0.08378564	0.07688031	0.08558512	0.09091042	0.08586309	0.08919872	0.08447325	0.09050078
GRUPO_4	0.10240649	0.08175308	0.08833021	0.08722378	0.08181455	0.07339337	0.07695854	0.08146111	0.08060055	0.08192212	0.08313612	0.08100009
GRUPO_5	0.07752215	0.07263517	0.08278848	0.08705063	0.09889987	0.08595720	0.07863790	0.08352487	0.08011068	0.08841184	0.08499766	0.07946355

FONTE: A Autora (2020)

O atributo Ocorrência\_Mes concentra a maior parte das probabilidades no grupo\_1 e grupo\_5, destacando janeiro no grupo\_1 com a maior probabilidade, de acordo com a Figura 33.

A Figura 34 apresenta o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Ocorrência\_Dia\_Semana.

FIGURA 34 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA\_DIA\_SEMANA NA BASE MANUAL

OCORRENCIA_DIA_SEMANA							
	SEGUNDA	TERÇA	QUARTA	QUINTA	SEXTA	SABADO	DOMINGO
GRUPO_1	0.16282743	0.14435523	0.15348164	0.17015489	0.14988373	0.12474222	0.09455487
GRUPO_2	0.13242153	0.14105448	0.14424635	0.15876691	0.15734462	0.13597724	0.13018887
GRUPO_3	0.13636563	0.14199816	0.14577268	0.15256097	0.15052741	0.14147148	0.13130367
GRUPO_4	0.09698190	0.09549129	0.10483450	0.10620217	0.12574915	0.22426776	0.24647325
GRUPO_5	0.12679356	0.13955772	0.14181153	0.14754647	0.15357151	0.15011269	0.14060652

FONTE: A Autora (2020)

Já o atributo Ocorrência\_Dia\_Semana concentra 4 das maiores probabilidades entre as 7 no grupo\_1 de acordo com a Figura 34. Percebe-se também que domingo contém a maior probabilidade com 0,24.

E a Figura 35 apresenta o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Ocorrência\_Hora.

FIGURA 35 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA\_HORA NA BASE MANUAL

OCORRENCIA_HORA				
	Manha	Tarde	Noite	Madrugada
GRUPO_1	0.33987100	0.34469747	0.27984731	0.03558422
GRUPO_2	0.27511329	0.35027950	0.28917077	0.08543644
GRUPO_3	0.23208930	0.33748336	0.33923895	0.09118839
GRUPO_4	0.12195347	0.20651873	0.37607954	0.29544826
GRUPO_5	0.21734764	0.37886328	0.28862161	0.11516747

FONTE: A Autora (2020)

O período da tarde no atributo Ocorrencia\_Hora foi a probabilidade de resultado mais alta com 0.37 no grupo\_5, neste atributo não teve um grupo dominante 4 foram redistribuídos, porém o grupo\_2 foi o único que teve menos resultados de acordo com a Figura 35.

Na Figura 36 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Regional\_Fato\_Nome.

FIGURA 36 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO REGIONAL\_FATO\_NOME NA BASE MANUAL

REGIONAL_FATO_NOME									
	BAIRRO_NOVO	BOA_VISTA	BOQUEIRAO	CAJURU	CIC	MATRIZ	PINHEIRINHO	PORTAO	REGIAO_METROPOLITANA
GRUPO_1	7.217761e-02	1.707692e-01	1.298758e-01	1.261902e-01	9.271204e-02	1.002589e-01	1.003905e-01	1.029792e-01	0.000000e+00
GRUPO_2	6.653326e-02	1.193729e-01	9.307710e-02	9.148943e-02	6.684748e-02	3.174511e-01	6.755863e-02	9.238250e-02	1.008831e-03
GRUPO_3	6.315743e-02	7.979167e-02	1.065352e-01	6.930201e-02	5.979255e-02	4.011382e-01	5.699823e-02	9.671851e-02	1.609293e-04
GRUPO_4	7.695854e-02	1.226450e-01	9.472293e-02	9.618281e-02	8.585610e-02	2.490703e-01	7.493008e-02	1.099364e-01	4.610136e-05
GRUPO_5	7.613862e-02	9.068797e-02	7.770067e-02	8.555553e-02	8.156115e-02	3.554995e-01	7.741057e-02	8.992926e-02	4.462991e-05

SANTA_FELICIDADE	
GRUPO_1	1.046466e-01
GRUPO_2	8.427877e-02
GRUPO_3	6.640528e-02
GRUPO_4	8.965178e-02
GRUPO_5	6.547207e-02

FONTE: A Autora (2020)

Por meio do resultado do algoritmo no atributo Regional\_Fato\_Nome apresentado na Figura 36, o grupo 4 concentra a maior parte das regionais que tiveram melhores probabilidades sendo as regionais (Bairro Novo, Boqueirão, Cajuru, Santa Felicidade e Região metropolitana), a regional portão no grupo 3 conteve a maior probabilidade em comparação com os resultados das regionais entre os grupos sendo (9,67).

FIGURA 37 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO FLAG\_FLAGRANTE NA BASE MANUAL

FLAG_FLAGRANTE		
	NAO	SIM
GRUPO_1	0.997893905	0.002106095
GRUPO_2	0.973456157	0.026543843
GRUPO_3	0.926718652	0.073281348
GRUPO_4	0.969788241	0.030211759
GRUPO_5	0.907281369	0.092718631

FONTE: A Autora (2020)

O resultado do atributo Flag\_Flagrante da Figura 37 tem a maior concentração das probabilidades na variável NÃO, porem está redistribuído os resultados entre os grupos.

Através da Figura 38 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Natureza\_Defesa\_Civil.

FIGURA 38 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO NATUREZA\_DEFESA\_CIVIL NA BASE MANUAL

NATUREZASIM_DEFESA_CIVIL		
	NAO	SIM
GRUPO_1	0.741125883	0.258874117
GRUPO_2	0.968891609	0.031108391
GRUPO_3	1.000000000	0.000000000
GRUPO_4	0.996373360	0.003626640
GRUPO_5	0.991721152	0.008278848

FONTE: A Autora (2020)

O atributo Natureza\_Defesa\_Civil tem a maior concentração das probabilidades na variável NÃO igual o atributo anterior, distribuído as probabilidades entre os grupos de acordo com a Figura 38.

Através da Figura 39 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados manual no atributo Origem\_chamado\_Descricao.

FIGURA 39 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ORIGEM\_CHAMADO\_DESCRICAO NA BASE MANUAL

ORIGEM_CHAMADO_DESCRICAO											
	Á_VIATURA	AOS_GMS	CASA_DA_MULHER_BRASILEIRA	CCO_BARIGUI	CCO_MUNICIPAL_URBS	CCO_PUC	CCO_RUAS_DA_CIDADANIA	CCO_SÍTIO_CERCADO			
GRUPO_1	1.377737e-02	2.904655e-02	0.000000e+00	3.510158e-04	0.000000e+00	4.387697e-05	1.140801e-03	4.387697e-05			
GRUPO_2	7.202395e-02	1.253432e-01	2.811497e-04	5.622995e-04	2.695730e-03	6.615288e-05	3.836867e-03	4.961466e-05			
GRUPO_3	7.313505e-02	9.265138e-02	7.314968e-05	5.705675e-04	1.462994e-05	1.170395e-04	2.253010e-03	1.609293e-04			
GRUPO_4	1.991579e-02	3.377693e-02	3.073424e-05	4.763807e-04	4.610136e-05	3.073424e-05	2.120663e-03	9.220272e-05			
GRUPO_5	4.166202e-02	7.004664e-02	3.347243e-04	1.785196e-03	6.694486e-05	4.462991e-05	3.146408e-03	1.338897e-04			
	CCO_URBS	CENTRAL_DE_ATENDIMENTO_PREFEITURA	CENTRAL_DE_FLAGRANTES	CIOSP	DEFESA_CIVIL	EMAIL	FONE	G5			
GRUPO_1	2.325479e-03	4.364442e-01	0.000000e+00	7.103681e-02	1.719977e-02	6.142776e-04	2.404458e-02	1.053047e-03			
GRUPO_2	1.003870e-02	1.931664e-02	1.653822e-05	4.200708e-02	4.084940e-03	1.488440e-03	6.881553e-02	8.881024e-03			
GRUPO_3	9.246119e-03	1.653183e-03	1.462994e-05	8.896464e-02	2.194490e-04	8.777961e-05	2.766521e-02	3.686744e-03			
GRUPO_4	8.129207e-03	2.985832e-02	0.000000e+00	1.452347e-01	7.683560e-04	1.997726e-04	2.595507e-02	4.920552e-02			
GRUPO_5	3.391873e-02	6.873006e-03	0.000000e+00	5.127976e-02	1.785196e-04	8.925981e-05	3.184344e-02	5.217236e-02			
	GUARDA_MUNICIPAL	IMEDIATA	NRGM	OFÍCIO	REDE_SOCIAL	SAMU	SIGA	TOTEM_RUA_SÃO_FRANSISCO	VEPPER		
GRUPO_1	3.341670e-01	6.678075e-02	1.755079e-04	1.667325e-03	8.775394e-05	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00		
GRUPO_2	2.649092e-01	3.042371e-01	8.765257e-04	2.804882e-02	1.653822e-04	4.152747e-02	5.953759e-04	8.269110e-05	4.961466e-05		
GRUPO_3	3.559610e-01	3.311486e-01	5.413076e-04	1.119190e-02	2.048191e-04	2.487089e-04	5.851974e-05	1.024096e-04	2.925987e-05		
GRUPO_4	6.122107e-01	6.646280e-02	1.690383e-04	3.857147e-03	2.612410e-04	1.229370e-04	8.605587e-04	1.075698e-04	1.075698e-04		
GRUPO_5	3.008502e-01	2.543905e-01	4.239841e-04	1.903466e-02	1.785196e-04	1.004173e-03	1.300515e-01	3.124093e-04	1.785196e-04		

FONTE: A Autora (2020)

O atributo Origem\_chamado\_Descricao contém 25 variáveis conforme a Figura 39, e percebe-se que o grupo predominante entre as probabilidades é o grupo\_2



contendo (10) variáveis, o grupo\_1 conteve (2) variáveis, o grupo\_3 (6) variáveis, o grupo\_4 (4) variáveis e o grupo\_5 (3) variáveis.

A Figura 40 contém os scripts utilizados para gerar a matriz confusão do algoritmo Naïve Bayes aplicados na base de dados Manual.

FIGURA 40 - SCRIPT DE GERAÇÃO DA MATRIZ CONFUSÃO DO NAÏVE BAYES

```
#Matriz confusão
matriz1 <- predict(Bayes, Manual)
table( Manual$CODIGODESCRICAO, matriz1)

#Matriz confusão
matriz2 <- predict(Bayes, K_means)
table( K_means$CODIGODESCRICAO, matriz2)
```

FONTE: A Autora (2020)

A função “predict” executada relaciona o resultado do Naïve Bayes com as bases de dados (Manual E k-Means). Através da segunda linha de comando do script “table” é relacionado o resultado da função do predict anterior cujo nome está matriz1 e 2 com o atributo meta “codigodescricao”, gerando a matriz confusão. A Figura 41 está apresentado a matriz confusão dos dados agrupados manual.

FIGURA 41 - MATRIZ CONFUSÃO DO NAÏVE BAYES DOS DADOS AGRUPADOS MANUAL

	matriz1				
	GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
GRUPO_1	13670	1917	2115	4890	199
GRUPO_2	2206	21206	22930	12281	1843
GRUPO_3	228	11449	37462	16803	2411
GRUPO_4	1708	5066	11689	44625	1986
GRUPO_5	663	7695	16426	10987	9042

FONTE: A Autora (2020)

De acordo com a matriz confusão da Figura 41 do algoritmo Naïve Bayes com os dados Manual apresentou uma taxa de acerto igual a

$$(13670 + 21206 + 37462 + 44625 + 9042/261497) = 48\% \text{ de taxa de acerto.}$$

A aplicação do algoritmo Naïve Bayes na base de Dados K-Means trouxe resultados em probabilidades, a primeira análise resultante do algoritmo é a distribuição dos grupos presente na Figura 42.

FIGURA 42 - DISTRIBUIÇÃO DE PROBABILIDADES POR GRUPO NA BASE K-MEANS

GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
0.01494090	0.04372899	0.12912194	0.54714968	0.26505849

FONTE: A Autora (2020)

Percebe-se por meio da Figura 42 que o grupo 4 teve o resultado maior comparado com os demais grupos.

E na Figura 43 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Atendimento\_Ano.

FIGURA 43 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ATENDIMENTO\_ANO NA BASE K-MEANS

ATENDIMENTO_ANO											
	A	B	C	D	E	F	G	H	I	J	K
GRUPO_1	0.10545175	0.09572562	0.09521372	0.08292808	0.09035065	0.07908881	0.09009470	0.06833888	0.08088047	0.09239826	0.11952905
GRUPO_2	0.13930914	0.07512024	0.08395278	0.08972453	0.08167906	0.08788806	0.09698295	0.07407084	0.08080455	0.09672059	0.09374727
GRUPO_3	0.08547312	0.06829557	0.11097290	0.06699245	0.07886865	0.07155338	0.08043832	0.07682511	0.08440693	0.10499037	0.17118318
GRUPO_4	0.08797299	0.08565957	0.06787906	0.07531556	0.09321489	0.10068634	0.09928151	0.08290583	0.08714128	0.08647032	0.13347265
GRUPO_5	0.10595568	0.09607283	0.09517832	0.07682652	0.10266620	0.10263735	0.10260849	0.07238285	0.07179132	0.08681036	0.08707006

FONTE: A Autora (2020)

Com o resultado do algoritmo no atributo Atendimento\_Ano da Figura 43 percebe-se uma maior concentração de melhor resultados das probabilidades no grupo\_5, sendo o ano (B, E, F, G ou 2010,2013,2014 e 2015).

Na Figura 44 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Ocorrencia\_Mes.

FIGURA 44 DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA\_MES NA BASE K-MEANS

OCORRENCIA MES											
	JANEIRO	FEVEREIRO	MARÇO	ABRIL	MAIO	JUNHO	JULHO	AGOSTO	SETEMBRO	OUTUBRO	NOVEMBRO
GRUPO_1	0.09393396	0.07908881	0.09444587	0.08932685	0.08292808	0.07320195	0.07678526	0.06833888	0.08036857	0.08907090	0.08599949
GRUPO_2	0.09129864	0.08587669	0.08613905	0.09567118	0.08027984	0.06462615	0.07284652	0.07844338	0.08150415	0.08841277	0.08727591
GRUPO_3	0.08067526	0.07673627	0.07827632	0.08239301	0.10848512	0.08334074	0.07119799	0.08117874	0.08630238	0.08976751	0.07655857
GRUPO_4	0.08255637	0.07665749	0.08400313	0.08483485	0.08287787	0.07623115	0.08320636	0.08909825	0.08352088	0.08634451	0.08398915
GRUPO_5	0.08617555	0.07440270	0.08618998	0.08340547	0.08842625	0.08036127	0.08235226	0.08460295	0.08187615	0.08630540	0.08388158

FONTE: A Autora (2020)

Segundo o resultado do atributo Ocorrencia\_Mes os grupos que tiveram maiores probabilidades é o grupo 3 e 4, destacando o mês de (maio) com o maior resultado conforme apresentado na Figura 44.

Através da Figura 45 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Ocorrendia\_Dia\_Semana.

FIGURA 45 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA\_DIA\_SEMANA NA BASE K-MEANS

OCORRENCIA_DIA_SEMANA							
	SEGUNDA	TERÇA	QUARTA	QUINTA	SEXTA	SABADO	DOMINGO
GRUPO_1	0.1313028	0.1394932	0.1604812	0.1430765	0.1540824	0.1471717	0.1243921
GRUPO_2	0.1245300	0.1383472	0.1334499	0.1526891	0.1463927	0.1580236	0.1465676
GRUPO_3	0.1278543	0.1289501	0.1343995	0.1682511	0.1542129	0.1447653	0.1415667
GRUPO_4	0.1239114	0.1262668	0.1321866	0.1362893	0.1457247	0.1661541	0.1694670
GRUPO_5	0.1305546	0.1362823	0.1407693	0.1434528	0.1435682	0.1588902	0.1464826

FONTE: A Autora (2020)

Em relação ao atributo Ocorrendia\_Dia\_Semana, o grupo\_1 é o que teve as maiores probabilidades, estando presente o dia da semana (Segunda, Terça e Quarta) conforme apresentado na Figura 45.

E a Figura 46 é apresenta o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Ocorrendia\_Hora.

FIGURA 46 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO OCORRENCIA\_HORA NA BASE K-MEANS

OCORRENCIA_HORA				
	Manha	Tarde	Noite	Madrugada
GRUPO_1	0.30765293	0.39083696	0.24827233	0.05323778
GRUPO_2	0.22544819	0.39562746	0.29514648	0.08377787
GRUPO_3	0.23245965	0.33496224	0.34660151	0.08597660
GRUPO_4	0.20108612	0.30671382	0.32270510	0.16949496
GRUPO_5	0.25278451	0.30695983	0.32083910	0.11941655

FONTE: A Autora (2020)

O período da tarde foi a probabilidade de resultado mais alto com (0.39) no grupo\_2 do atributo Ocorrendia\_Hora, não teve um grupo dominante cada período ficou distribuído em um grupo, porém o grupo\_5 foi o único que teve menos resultados de acordo com a Figura 46.

Na Figura 47 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Regional\_Fato\_Nome.



FIGURA 47 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO REGIONAL\_FATO\_NOME NA BASE K-MEANS

REGIONAL_FATO_NOME									
	BAIRRO_NOVO	BOA_VISTA	BOQUEIRAO	CAJURU	CIC	MATRIZ	PINHEIRINHO	PORTAO	REGIAO_METROPOLITANA
GRUPO_1	0.0621960584	0.1159457384	0.0962375224	0.0836959304	0.0708983875	0.3235218838	0.0688507807	0.0959815715	0.0000000000
GRUPO_2	0.0718845649	0.1230432882	0.1185832969	0.0763445562	0.0641014429	0.2819414080	0.0618277219	0.1032794053	0.0011368605
GRUPO_3	0.0659854879	0.1307567007	0.0939138161	0.1128979713	0.0690656005	0.2868650970	0.0739523175	0.0858581371	0.0001480823
GRUPO_4	0.0670962692	0.1025244971	0.0917960833	0.0763918981	0.0661317603	0.3608101875	0.0611414753	0.0961223948	0.0003564489
GRUPO_5	0.0795244691	0.1105869114	0.1079033934	0.1052631579	0.0963758079	0.2198753463	0.0922351108	0.1084949215	0.0001154201

SANTA_FELICIDADE	
GRUPO_1	0.0826721270
GRUPO_2	0.0978574552
GRUPO_3	0.0805567896
GRUPO_4	0.0776289856
GRUPO_5	0.0796254617

FONTE: A Autora (2020)

O Resultado do algoritmo no atributo Regional\_Fato\_Nome da Figura 47 concentra a maior parte das regionais no grupo\_2 que tiveram melhores probabilidades sendo (Boqueirão, Portão, Santa Felicidade e Região metropolitana), a regional matriz no grupo 4 conteve a maior probabilidade em comparação com os resultados das regionais entre os grupos sendo (0,36).

Através da Figura 48 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Flag\_flagrante.

FIGURA 48 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO FLAG\_FLAGRANTE NA BASE K-MEANS

FLAG_FLAGRANTE		
	NAO	SIM
GRUPO_1	0.92398260	0.07601740
GRUPO_2	0.95863577	0.04136423
GRUPO_3	0.92477417	0.07522583
GRUPO_4	0.95480088	0.04519912
GRUPO_5	0.95663089	0.04336911

FONTE: A Autora (2020)

A maior concentração das probabilidades do atributo Flag\_flagrante está na instância NÃO e os resultados estão distribuídos nos grupos de forma uniforme.

Na Figura 49 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Natureza\_Defesa\_Civil.

FIGURA 49 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO NATUREZA\_DEFESA\_CIVIL NA BASE K-MEANS

	NATUREZA_NATUREZA_CIVIL	
	NAO	SIM
GRUPO_1	0.7315076	0.2684924
GRUPO_2	0.7648448	0.2351552
GRUPO_3	0.8622834	0.1377166
GRUPO_4	1.0000000	0.0000000
GRUPO_5	1.0000000	0.0000000

FONTE: A Autora (2020)

A maior concentração das probabilidades no atributo Natureza\_Defesa\_Civil está na variável NÃO igual do atributo anterior, porém as maiores probabilidades estão do grupo 4 e 5.

Através da Figura 50 é apresentado o resultado da probabilidade do algoritmo aplicado na base de dados agrupada pelo K-Means no atributo Origem\_chamado\_Descricao.

FIGURA 50 - DISTRIBUIÇÃO DE PROBABILIDADES DO ATRIBUTO ORIGEM\_CHAMADO\_DESCRICAO NA BASE K

	ORIGEM_CHAMADO_DESCRICAO														
	À_VIATURA	AOS_GMS	CASA_DA_MULHER_BRASILEIRA	CCO_BARIGUI	CCO_MUNICIPAL_URBS	CCO_PUC	CCO_RUAS_DA_CIDADANIA	CCO_SÍTIO_CERCADO							
GRUPO_1	5.502943e-02	1.226005e-01	0.000000e+00	0.000000e+00	5.119017e-04	0.000000e+00	3.071410e-03	0.000000e+00							
GRUPO_2	4.967206e-02	1.151727e-01	0.000000e+00	7.870573e-04	0.000000e+00	8.745081e-05	2.448623e-03	8.745081e-05							
GRUPO_3	5.490893e-02	9.083370e-02	5.923293e-04	7.107952e-04	1.480823e-04	5.923293e-05	3.080113e-03	5.923293e-05							
GRUPO_4	5.182488e-02	7.789457e-02	1.048379e-04	9.155845e-04	5.102112e-04	9.085953e-05	2.753743e-03	9.085953e-05							
GRUPO_5	4.009407e-02	5.631060e-02	5.771006e-05	4.039705e-04	1.298476e-03	1.442752e-05	2.207410e-03	1.587027e-04							
	ORIGEM_CHAMADO_DESCRICAO														
	CCO_URBS	CENTRAL_DE_ATENDIMENTO_PREFEITURA	CENTRAL_DE_FLAGRANTES	CIOSP	DEFESA_CIVIL	EMAIL	FONE	GG5							
GRUPO_1	1.049399e-02	1.126184e-02	0.000000e+00	1.023803e-01	2.124392e-02	2.303558e-03	6.961863e-02	3.327361e-03							
GRUPO_2	1.469174e-02	1.180586e-02	0.000000e+00	9.934412e-02	1.757761e-02	1.049410e-03	4.083953e-02	1.923918e-03							
GRUPO_3	1.098771e-02	6.195765e-02	0.000000e+00	6.536354e-02	1.060270e-02	2.369317e-04	4.418777e-02	5.479046e-03							
GRUPO_4	1.552999e-02	8.477893e-03	6.989195e-06	1.001132e-01	3.494597e-04	5.940815e-04	3.862928e-02	2.319015e-02							
GRUPO_5	7.776431e-03	1.442030e-01	1.442752e-05	5.658472e-02	2.885503e-04	1.875577e-04	2.824908e-02	4.061346e-02							
	ORIGEM_CHAMADO_DESCRICAO														
	GUARDA_MUNICIPAL	IMEDIATA	NRGM	OFÍCIO	REDE_SOCIAL	SAMU	SIGA	TOTEM_RUA_SÃO_FRANCISCO	VEPPER						
GRUPO_1	3.140517e-01	2.674686e-01	7.678526e-04	1.407730e-02	1.023803e-03	5.119017e-04	2.559509e-04	0.000000e+00	0.000000e+00						
GRUPO_2	3.474421e-01	2.808045e-01	6.996065e-04	1.469174e-02	3.498032e-04	2.623524e-04	8.745081e-05	1.749016e-04	0.000000e+00						
GRUPO_3	3.853398e-01	2.230712e-01	9.477269e-04	3.912335e-02	2.961647e-04	9.773434e-04	9.181105e-04	8.884940e-05	2.961647e-05						
GRUPO_4	4.041292e-01	2.421337e-01	3.844057e-04	1.388054e-02	1.327947e-04	1.759879e-02	4.193517e-04	1.747299e-04	6.989195e-05						
GRUPO_5	3.643381e-01	1.709084e-01	3.751154e-04	1.053209e-03	2.019852e-04	3.606879e-04	8.412685e-02	4.328255e-05	1.298476e-04						

FONTE: A Autora (2020)

O atributo Origem\_chamado\_Descricao de acordo com a Figura 50, percebe-se que o grupo\_1 conteve (3) Instância, o grupo\_2 (3) Instância, o grupo\_3(9) Instância, o grupo\_4 (9) Instância e o grupo\_5 (1) Instância. Portanto nota se que o grupo 3 e 4 foram os que tiveram as melhores probabilidades entre os grupos e 25 variáveis.

A Figura 51 está apresentando a matriz confusão dos dados agrupados pelo K-means.

FIGURA 51 - MATRIZ CONFUSÃO DO NÁIVE BAYES DOS DADOS AGRUPADOS KMEANS

	matriz2				
	GRUPO_1	GRUPO_2	GRUPO_3	GRUPO_4	GRUPO_5
GRUPO_1	4	268	787	2821	27
GRUPO_2	6	770	1918	8682	59
GRUPO_3	2	418	4791	28304	250
GRUPO_4	0	7	308	139715	3048
GRUPO_5	0	0	49	51893	17370

FONTE: A Autora (2020)

De acordo com a matriz confusão da Figura 51 do algoritmo Naïve Bayes com os dados K-Means apresentou uma taxa de acerto igual a

$$(4 + 770 + 4791 + 139715 + 17370 / 261497) = 62\% \text{ de taxa de acerto.}$$

#### 4.4 PÓS PROCESSAMENTO

Por meio das análises realizadas neste trabalho na base da Guarda Municipal de Curitiba, o presente capítulo de pós-processamento apresentou os resultados alcançados com este estudo através do pré-processamento, estatística e mineração de dados.

No capítulo de pré-processamento foram realizadas limpezas na base de dados e o tratamentos. A base de dados da Guarda Municipal ao todo apresentou (35) atributos e (280.905) registros de ocorrências entre o ano de 2009 até o mês 5 de 2020, porém através da limpeza de dados (Valores Ruidosos, Valores Ausentes, Valores Inconsistentes, Valores Redundantes) foi reduzido para (261.493) ocorrências e (13) atributos e com o período de janeiro de 2019 até dezembro de 2019.

Na aplicação de estatísticas no capítulo 4.2 foi possível identificar tendencias e padrões na base de dados. Foi analisado que a maior concentração das ocorrências está no ano de 2019 com (32.451) registros e uma média de (23.772) ocorrências atendidas por ano. A principal origem das ocorrências é pelo 153 - telefone da Guarda Municipal através de (101.286) ocorrências registradas, analisando entre os anos foi possível identificar uma tendencia de que está aumentando cada vez mais o número de ocorrências sendo registradas através do telefone 153. A principal secretaria responsável pelos atendimentos é a SMDT- Secretaria Municipal de Defesa Social e

Trânsito com (137.718) ocorrências somadas entre os anos, e foi observado também uma tendencia de aumento de ocorrência por período de anos. A maior parte das ocorrências tem uma tendencia em não ter flagrante, somando todos anos ao todo foi atendido (248.714) ocorrências. A maior parte das ocorrências também não teve distribuição de defesa civil, com (253.109) registros somando os períodos manteve se uma tendencia na base de dados. Entre as 167 descrições o apoio é a mais tendenciosa ocorrida na base com um total de (39.733) ocorrências atendidas somando os períodos. Foi observado que o atributo Ocorrencia\_Dia\_Semana sábado e domingo são os dias em que mais tem as ocorrências com uma média em ambos de (41mil) somando os anos, e tem se mantido uma média de (3.700) ocorrências por ano no sábado e domingo. Os meses de janeiro a dezembro manteve uma padrão em quantidade de atendimento de ocorrências somando os anos de (21.791) atendimentos. O período de atendimentos tarde e noite é um padrão identificado na base, é apresentado uma média de (80 mil) ocorrências somadas entre os anos e uma média de (7.500) ocorrência por mês entre os dois maiores períodos. Com a análise das regionais foi possível identificar uma tendencia, a regional Matriz contém a maior quantidade de atendimentos com um total de (81.038) ocorrências somadas entre os anos, porém por ela concentrar a maior parte dos bairros acaba se tornando uma tendencia em ser a mais atendida pela Guarda, principalmente por conter também o bairro Centro que ao todo somando entre os anos atendeu (44.160) ocorrências e é onde concentra a maior quantidade de Guardas Municipais.

Durante o pré-processamento no tratamento das bases de dados foi realizado o agrupamento do atributo natureza1\_descricao, este atributo continha 167 instâncias únicas e para a aplicação dos algoritmos de mineração de dados o processamento poderia levar até dias calculando tudo. Para a realização dos agrupamentos foi considerado duas metodologias diferentes, e elas têm como principal diferença o método de agrupamento realizado no campo natureza1\_descricao, sendo um realizado de modo manual classificando em 5 grupos conforme apresentado a redistribuição no apêndice K, já o segundo agrupamento foi realizado através do algoritmo K-Means agrupando em 5 grupos com base nos próprios dados da base disponibilizado no apêndice L.

Após a realização dos agrupamentos foram selecionados os algoritmos para aplicação da mineração dos dados. Por se tratar de uma base de dados qualitativa foram selecionados a árvore de decisão e o Naïve Bayes. Foi realizada a aplicação

de cada algoritmo duas vezes considerando as duas bases de dados com agrupamentos diferentes a Manual e K-Means

Em comparação do Naïve Bayes e da Árvore de Decisão a taxa de acerto foi 2% melhor no algoritmo Naïve Bayes. A base de dados agrupados pelo

k-Means teve em torno de 60% de taxa de acerto, já na base de dados manual obteve-se em torno de 40%, apesar do resultado ter sido maior na base do K-Means os dados foram agrupados de forma automática obtendo um agrupamento disperso com as instâncias de forma aleatória do atributo Natureza\_descrição. Cada cluster não seguiu um padrão fazendo com que a taxa de acerto da base de dados pelo K-Means ficasse melhor por concentrar a maior parte das instâncias apenas em um grupo dos clusters.

#### 4.4.1 Resultados do algoritmo árvore de decisão

Através das bases de dados Manual e K-mens foi realizado a aplicação do algoritmo de Árvore de Decisão. Na primeira árvore gerada conforme a FIGURA 24 do capítulo de mineração de dados, com os dados agrupados manualmente obteve-se uma taxa de acerto igual a 46% ( $(14099 + 17545 + 47589 + 36145 + 5828)/261497 = 46\%$ ). Já a segunda árvore da FIGURA 25 do capítulo de mineração de dados, com os dados agrupados pelo k-Means obteve-se 62% de taxa de acerto ( $(0 + 0 + 4650 + 141805 + 15826)/261497 = 62\%$ ) conforme apresentado nas matrizes confusão da FIGURA 27 com os dados agrupados manualmente e a Figura 28 com os dados agrupados pelo K-Means.

Por meio das análises das bases de dados foi possível observar diferenças como: o tamanho da árvore de decisão com os dados Manual é maior (em número de folhas e profundidade) quando comparada à aplicada na base de dados pelo K-Means. Na base manual ficaram mais redistribuídas as ramificações e o nó raiz foi o Grupo 3 a qual contém o agrupamento pelas descrições relacionadas ao Dano disponibilizado no apêndice K. Já a árvore gerada pelo K-Means apresenta no nó raiz o Grupo 4.

#### 4.4.2 Resultados do algoritmo Naïve Bayes

O Naïve Bayes foi o segundo algoritmo analisado neste trabalho e a partir dele obteve-se uma taxa de acerto parecida com algoritmo Árvore de decisão. De acordo com a matriz confusão da Figura 41 do algoritmo Naïve Bayes com os dados agrupados Manual apresentou uma taxa de acerto igual a 48% ( $13670 + 21203 + 37462 + 44626 + 9041 / 261497 = 48\%$ ). Já a matriz confusão aplicada a base de dados agrupada pelo K-Menos da Figura 51 obteve 62% de taxa de acerto ( $4 + 770 + 4791 + 139714 + 17369 / 261497 = 62\%$ ).

A análises dos resultados pelo algoritmo Naïve Bayes, foi observado com a aplicação na base de dados Manual, que o grupo que teve a maior probabilidade de atendimento das ocorrências foi o grupo 3. Ainda viabilizou análises relevantes tais como: os anos de 2009 até 2013 (A ate E) concentram- se no grupo\_1 e os anos de 2015 até 2017e 2019 (G, H, I, K) concentram-se no Grupo\_3. Já na base de dados agrupadas pelo K-Means observa-se que a maior probabilidade de ocorrências atendidas é no grupo 4 e concentra-se a maior parte das probabilidades no atributo atendimento ano no grupo 5 sendo o ano (B, E, F, G ou 2010,2013,2014 e 2015).

## 5 CONSIDERAÇÕES FINAIS

A cada dia que se passa vem aumentando o número de dados públicos, nas organizações, na internet, em aplicativos e outros meios através dos avanços tecnológicos, e com esta massiva quantidade de dados cada vez mais vem sendo procurada a área de análise dos dados. Com a Lei Geral de Proteção de Dados (LGPD), em vigor desde o dia 18 de setembro de 2020, a extração e análise de muitas informações que geralmente são dados pessoais apresenta novos desafios legais, apesar disso o mercado de análise de dados nunca esteve tão aquecido, e os analistas estão se adaptando a este novo cenário.

A mineração de dados vem sendo muito utilizada pelos analistas tanto de forma isolada quanto complementar às análises estatísticas. Pensando neste cenário em se analisar dados e utilizar técnicas de análise surgiu-se o interesse de desenvolver o presente trabalho.

O presente trabalho a fim de responder o problema de pesquisa “Como prever tendências das ocorrências atendidas na Guarda Municipal de Curitiba?” se dá através da aplicação do KDD (Knowledge-Discovery in Databases) a qual é uma metodologia que se inicia na obtenção dos dados, pré processamento, mineração de dados e o pós processamento. Por meio de análises realizadas em tabelas dinâmicas no Excel e algoritmos aplicados no R Studio foi obtida a resposta à pergunta.

### 5.1 ALCANCE DOS OBJETIVOS

O objetivo geral deste trabalho concentrou-se na identificação de tendências e padrões que possibilitem a descoberta de conhecimento na base da Guarda Municipal de Curitiba. Este objetivo foi alcançado e apresentado no capítulo de pós processamento (seção 4.4). Nesta seção são apresentados resultados de tendências e padrões como, por exemplo, a regional Matriz apresentar a maior parte das ocorrências atendidas. Observou-se que os meses entre janeiro e dezembro mantiveram um padrão de ocorrências registradas com uma média de (21.791) registros somando os anos.

O primeiro objetivo específico foi a realização de estudos e pesquisa com levantamento bibliográfico. Através de pesquisas em bibliotecas virtuais da (EBSCOHOST, CAPES e SciELO) foi obtido o total de teses em três análises de

termos apresentadas no capítulo de justificativa. Através dos termos ("Guarda Municipal; "Guarda Municipal" AND "Análise"; "Guarda Municipal" AND "Dados") foi identificado os três termos não eram muito estudados, e existem poucos artigos e teses que falavam sobre o assunto. Por meio desta análise identifica-se a importância no desenvolvimento deste trabalho conforme descrito na justificativa,

Estudar a base e selecionar os principais atributos a serem analisados foi o segundo objetivo específico definido. Este objetivo foi alcançado e resultou em uma base de dados tratada no capítulo de pré processamento, a base inicialmente continha 35 atributos e 280.905 ocorrências com dados de 2019 até o mês 5 de 2020, seguindo as análises e tratamentos foi reduzido para 261.493 ocorrências e 13 atributos com o período de janeiro de 2009 até dezembro de 2019. Para aplicação dos algoritmos optou-se por trabalhar com duas bases de dados, a decisão em usar duas bases surgiu pelo fato do atributo natureza1\_descricao continha 167 instâncias únicas, e para a aplicação dos algoritmos de mineração de dados o processamento poderia levar horas até dias calculando tudo, portanto foi necessário realizar agrupamentos nas descrições das ocorrências atendidas, resultando em uma base de dados agrupada manualmente e uma base de dados agrupada pelo K-Means. Para a realização dos agrupamentos foi considerado duas metodologias diferentes, a primeira realiza o agrupamento de forma manual conforme apresentado a redistribuição no apêndice K gerando 5 grupos, já o segundo agrupamento foi realizado através do algoritmo K-Means no RStudio agrupando em 5 grupos classificados de acordo com o apêndice L.

O terceiro objetivo específico foi calcular estatísticas descritivas para análise dos dados na base de dados. A base de dados da Guarda Municipal de Curitiba contém todos os dados qualitativos repartidos entre nominal e ordinal, portanto, por seguir este tipo de dado foi realizado análises com tabelas de frequências absolutas, gráficos de barras, média, mínimo e máximo.

O quarto objetivo concentrou-se na aplicação dos algoritmos de mineração de dados, foram selecionados os algoritmos Árvore de decisão e Naïve Bayes considerando que a base de dados contém somente dados qualitativos, e para aplicação do algoritmo foi executado duas vezes cada algoritmo pelo fato de conter duas bases de dados, sendo uma base com os dados agrupados no atributo natureza1\_descricao de forma manual e a segunda base com os dados agrupados pelo k-Means.



O quinto e último objetivo específico, foi a análise dos resultados, realizando um comparativo entre os resultados do método Rpart e o Naïve Bayes com a matriz confusão, taxa de acerto e análise de tendências e padrões. Percebeu-se que a base de dados com agrupamento k-Means obtinha um resultado melhor comparando o mesmo algoritmo com a base de dados agrupadas manualmente, porém a taxa de acerto na base de dados K-Means está fortemente ligada a concentração da maior parte dos resultados no grupo 4 levando a taxa de acertos a um resultado maior que o da base manual. Em comparação entre os dois algoritmos o Naïve Bayes obteve-se também uma taxa de acerto 2% maior que a árvore do método Rpart de decisão. O algoritmo Naïve Bayes contribuiu para as análises dos atributos uma vez que, por meio das probabilidades geradas, foi identificado que o atributo Regional na base de dados manual concentra a maior parte dos atendimentos no grupo 4 com as regionais Bairro Novo, Boqueirão, Cajuru, Santa Felicidade e Região metropolitana. Na base de dados pelo K-Means a maior concentração dos atendimentos está no grupo 2 com regionais Boqueirão, Portão, Santa Felicidade e Região metropolitana. Com a árvore de decisão foi possível identificar a ramificação gerada pelo algoritmo e pelas análises complementares apresentada na seção 4.4 foi observado que através de uma análise simples de estatística foi possível identificar tendências e padrões.

Os objetivos específicos e geral propostos para elaboração deste trabalho foram atingidos, foi realizado pesquisas e análises identificando padrões e regras de classificação e comparativo entre os resultados possibilitando descoberta de conhecimentos.

## 5.2 CONTRIBUIÇÕES

Através do presente trabalho foi possível identificar a falta de pesquisas em análise de dados da Guarda Municipal de Curitiba e sua elaboração contribuiu com a geração de conhecimento em um tema tão importante que é a Segurança Pública. A segunda contribuição deste trabalho foi para a área de análise de dados e mineração de dados. Por ser um trabalho que analisou dados públicos, pode incentivar outros pesquisadores a utilizar os dados abertos. Este trabalho contribui como suporte a tomada de decisão na gestão urbana através do uso dos resultados, e estes

resultados podem estar auxiliando a identificar o bairro, hora, dia da semana em que mais ocorre ocorrências para predição. E a principal contribuição foi para o meu conhecimento e desenvolvimento profissional na área de dados utilizando o aprendizado para ser aplicado em outros contextos.

### 5.3 TRABALHOS FUTUROS

Apesar de haver poucos artigos sobre análise de dados da Guarda Municipal o tema segurança pública sempre foi algo muito discutido, e seguindo o âmbito de trabalhos futuros sugere-se a realização de comparações com outras bases de dados da área da segurança, por exemplo da polícia militar ou civil.

Sugere-se também a aplicação destas análises colocadas em um Power Bi com mapas interativos apresentando toda a análise redistribuído entre os locais de ocorrências.

Por fim recomenda-se também a utilização de outros tipos de métodos de análises e ferramentas, comparando o resultado obtido em ferramentas diferentes.

## REFERÊNCIAS

ASSIS, Wilson Martins de. **Gestão da informação nas organizações**: Como analisar e transformar em conhecimento informações captadas no ambiente de negócios. Autêntica, 2008. 140 p.

BARBOSA, Denise Chaves Carvalho. **Mineração de dados usando o software WizRule em base de dados de compras de itens de TI**. 2007. 80 f. Dissertação (Mestrado) - Curso de Administração, Faculdade de Economia e Finanças Ibmecc, Rio de Janeiro, 2007.

BOENTE, Alfredo Nazareno Pereira; OLIVEIRA, Fabiano Saldanha Gomes de; ROSA, José Luiz dos Anjos. **Utilização de ferramentas de kdd para integração de aprendizagem e tecnologia em busca da gestão estratégica do conhecimento na empresa**. 12 f. Monografia (Especialização) - Centro Universitário Estadual da Zona Oeste, Rio de Janeiro.

BERTONCINI, Cristine; BRITO, Asriana; SILVA, Ismael; LEME, Elisangela. **Processo decisório: a tomada de decisão**. São Paulo: Faeg/aceg, 12 p.

BONASSA, Mariana Dandolini. **Processo de recuperação da informação: habilidades e competências do bibliotecário**. 2009. 103 f. TCC (Graduação) - Curso de Biblioteconomia, Universidade Federal de Santa Catarina, Florianópolis.

CARAVANTES, Geraldo; PANNO, Cláudia; KLOECKNER, Mônica. **Administração: Teorias e processo**. São Paulo: Pearson, 2005. 572 p.

CARVALHO, Claudio Frederico de. **Guarda municipal**: O que você precisa saber sobre guarda municipal e nunca teve a quem perguntar. 3. ed. Curitiba: do Autor, 2011. 284 p.

CAMPOS, William Sergio Antunes de. **A guarda municipal no contexto da segurança pública integrada**: desafios e perspectivas no exercício funcional frente à demanda por segurança e proteção do cidadão. 2013. 102 f. Tese (Mestrado) -

Curso de Gestão Pública, Centro de Ciências Jurídicas e Econômicas, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

CARDOSO, Olinda Nogueira Paes; CARDOSO, Olinda Nogueira Paes. **Recuperação de informação**. Lavras, 2004. 6 p.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados**: Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva Educação, 2016. 559 p.

CHIAVENATO, Idalberto. **Introdução à teoria geral da administração**. 7. ed. Rio de Janeiro: Campus, 2003. 630 p.

CHOO, Chun Wei. **A organização do conhecimento**: Como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. Tradução Eliana Rocha. - São Paulo: Editora Senac São Paulo, 2003.,

DADOS ABERTOS. **Consulta de bases** Disponível em: <https://www.curitiba.pr.gov.br/dadosabertos/busca/> Acesso em: 25 set. 2020.

DAVENPORT, Thomas H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998. 312 p. (ISBN 85-86082-72-4) Tradução Bernadette Siqueira Abrão

NÓBREGA, Emeide Duarte; BASTOS, Simone Paiva; ARAUJO, Alzira Karla Silva. **Múltiplas abordagens da gestão da informação e do conhecimento no contexto acadêmico da ciência da informação**. João Pessoa: Editora da UFPB, 2014. 168 p.

FAZ.Serviços Disponível em: <https://fas.curitiba.pr.gov.br/maparegional.aspx> Acesso em: 25 set. 2020.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery in databases**. AI magazine, v. 17, n. 3, p. 37, 1996. Disponível em: <<http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>>. Acesso em: 22 mar. 2019.

FIGUEREIDO, Dayana Ester Andrade. **Recuperação da informação**: Uma análise sobre os sistemas de busca da web. 2006. 64 f. Tese (Graduação) - Curso de Biblioteconomia, Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas S.a, 2008. 220 p.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data mining**: Um guia prático. 4. ed. Rio de Janeiro: Elsevier, 2005.

GROHS, Regina Cecilia. **A estatística como fonte explicativa do trabalho de defesa social**: Guarda municipal de Curitiba. Curitiba, 2003. 140 p.

MARCHIORI, Patrícia Zeni. **A ciência e a gestão da informação**: Compatibilidades no espaço profissional. 2002. 79 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Federal do Paraná, Curitiba, 2002.

MAXIMIANO, A. C. A. **Introdução à administração**. 5. ed. rev. e ampl. – São Paulo:Atlas, 2000.

MELLO, Luis Cesar de. **Um assistente de feedback para o serviço de filtragem do software direto**. 115 f. Dissertação (Mestrado) - Curso de Instituto de Informática Programa de Pós-Graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

MILANI, Cristian Simioni; CARVALHO, Deborah Ribeiro. **Pós-processamento em kdd**. Curitiba: Revista de Engenharia e Tecnologia, 2013.

MORAES, G. D. A.; FILHO, E. E. **A gestão da informação diante das especificidades das pequenas empresas**. *Ci. Inf.*, Brasília, v. 35, n. 3, p. 124-132, set./dez. 2006. Disponível em: <http://www.scielo.br/pdf/ci/v35n3/v35n3a12.pdf>. Acesso em: 11 mai. 2019.

MOREIRA, Cleverson Bayer. **Gestão da informação**. Guarapuava: Gráfica Unicentro, 2014.

NEVES, Rita de CÁssia David das. **Pré-processamento no processo de descoberta de conhecimento em banco de dados**. 137 f. Tese (Mestrado) - Curso de Ciência da Computação, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

PINTO, Cláudio Alexandre Morais; VIEIRA, Hugo Abel da Silva; BULHÕES, Luís Miguel Parreira. **Data mining em R**. 2005. 41 f. TCC (Graduação) - Curso de Licenciatura em Informática de Gestão, Universidade do Minho, Azurém, 2005.

PONTES JUNIOR, João de; CARVALHO, Rodrigo Aquino de; AZEVEDO, Alexander William. **Recuperação da informação à recuperação do conhecimento: Reflexões e propostas**. 4. ed.: Perspectivas em Ciência da Informação, 2013. 16 p.

GIACOMITTI, Regina Cecília Gröhs A estatística como fonte explicativa do trabalho da Defesa Social da Guarda Urbana de Curitiba. 2003. Monografia (Especialização Em Controle Est. de Qualidade) - Universidade Federal do Paraná

REIS, Edna Afonso; REIS, Ilka Afonso. **Análise descritiva de dados**. 64 f. Dissertação (Mestrado) - Curso de Instituto de Ciências Exatas Departamento de Estatística, Universidade Federal de Minas Gerais, Minas Gerais, 2002.

REZENDE, Y. **Informação para negócios: Os novos agentes do conhecimento e a gestão do capital intelectual**. Ciência da Informação, Brasília, v. 31, n. 2, p. 120-128, maio/ago. 2002.

SAYÃO, Luis Fernando; SALES, Luana Farias. CURADORIA DIGITAL: um novo patamar para preservação de dados digitais de pesquisa. **Curadoria digital**, João Pessoa, v. 22, n. 3, p. 179-191, 01 dez. 2012.

SILVA, Edna Lucia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: Ufsc, 2005. 139 p.

SILVA, Leandro Augusto da; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução a mineração a mineração de dados com aplicações em R**. 6. ed. São Paulo: Elsevier, 2016. 277 p.

SILVA, Mayane Paulino de Brito e; VECHIATO, Fernando Luiz. **Usabilidade na recuperação da informação em catálogos bibliográficos**: Avaliação do catálogo online do SIGAA na perspectiva dos pós-graduandos da UFRN. **Inf. & Soc.:est**, João Pessoa, v. 27, n. 3, p.277-289, 10 out. 2017.

SOARES, Fabiana de MENEZES; JARDIM, Tarciso Dal Maso; HERMONT, Thiago Brasileiro Vilar. **Lei de acesso à informação**: Acesso à informação. Brasília: Senado Federal, 2013. 44 p. Disponível em: <https://www2.senado.leg.br/bdsf/item/id/496325>. Acesso em: 11 mai. 2020.

SENADO FEDERAL. **Lei de acesso à informação no Brasil**: O que você precisa saber. Brasília-df:SEEP/Senado Federal, 2013. v. 1, 44 p.

TARAPANOFF, Kira. **Inteligência, informação e conhecimento em corporações**. Brasília: IBICT, UNESCO, 2006. 456 p. Universidade de São Paulo, SP, 2000.

## APÊNDICE

### APÊNDICE A - ATRIBUTOS E INSTÂNCIAS DA BASE DE DADOS DAS OCORRÊNCIAS DA GUARDA MUNICIPAL DE CURITIBA

Atributo	Valor do atributo	Tipo de atributo	Descrição
ATENDIMENTO_ANO	Variável Ex: 2009	Inteiro	Ano da ocorrência Realizada
ATENDIMENTO_BAIRRO_NOME	Variável: ABRANCHES ÁGUA VERDE AHÚ ALTO BOQUEIRÃO ALTO DA GLÓRIA ALTO DA RUA XV ATUBA AUGUSTA BACACHERI BAIRRO ALTO BAIRRO FICTÍCIO BAIRRO NAO INFORMADO BAIRRO NÃO LOCALIZADO BARREIRINHA BATEL BIGORRILHO BOA VISTA BOM RETIRO BOQUEIRÃO BORDA DO CAMPO BRAGA BUTIATUVINHA CABRAL CACHOEIRA CAJURU CAMPINA DO SIQUEIRA CAMPO COMPRIDO CAMPO DE SANTANA CAMPO DE SÃO BENEDIT CAMPO PEQUENO CAMPO PEQUENO CANGUIRI CAPÃO DA IMBUIA CAPÃO RASO CASCATINHA CAXIMBA CENTRO CENTRO CÍVICO CIDADE INDUSTRIAL	Nominal	Nome do bairro em que foi realizado o atendimento



	CIDADE JARDIM COLOMBO CRISTO REI FANNY FAZENDINHA FERRARIA FORMOSO GANCHINHO GRALHA AZUL GUABIROTUBA GUÁIRA HAUER HUGO LANGE IGUAÇÚ 1 INDICAÇÕES CANCELADA JARDIM BOA VISTA JARDIM BOM PASTOR JARDIM BOTÂNICO JARDIM COLONIAL JARDIM DAS AMÉRICAS JARDIM INDUSTRIAL JARDIM LOANDA JARDIM SOCIAL JARDIM WEISSOPOLIS JD EUROPA JD IPE JD SUISSA JUVEVÊ LAMENHA PEQUENA LINDÓIA LOT. MARINONI MARACANÃ MARIA ANTONIETA MENINO DEUS MERCÊS MONTE REY MOSSUNGUÊ NÃO ENCONTRADO NÃO INFORMADO NF NI NOVO MUNDO ORLEANS PAROLIN PILARZINHO PINHAIS PINHEIRINHO PLANTA MEIRELES PORTÃO PRADO VELHO QUATRO BARRAS REBOUÇAS RIO VERDE RIVIERA ROÇA NEGRA SANTA CÂNDIDA SANTA FELICIDADE		
--	---	--	--

	SANTA QUITÉRIA SANTA TERESINHA SANTA TEREZINHA SANTO ANTONIO SANTO INÁCIO SÃO BRAZ SÃO CRISTOVÃO SÃO FRANCISCO SAO GERONIMO SÃO JOÃO SÃO JORGE SÃO JOSE SAO JOSE DOS PINHAIS SÃO JUDAS TADEU SÃO LOURENÇO SÃO MIGUEL SÃO PEDRO SÃO THOMAS SÃO THOMAZ SEM DADOS SEMINÁRIO SÍTIO CERCADO SITIO DAS PALMEIRAS TABOÃO TANGUA TARUMÃ TATUQUARA THOMAS COELHO TINGUI TINGÜI UBERABA UMBARÁ VENEZA VILA BANCÁRIA VILA FORMOSA VILA IZABEL VILA MARIA ANTONIETA VILA PERNETA VISTA ALEGRE XAXIM		
FLAG_EQUIPAMENTO_URBANO	Variável: SIM NÃO	Nominal	Flag para identificar se é um equipamento urbano ou não
FLAG_FLAGRANTE	Variável: SIM NÃO	Nominal	Flag para identificar se é um equipamento urbano ou não
NATUREZA1_DEFESA_CIVIL	Variável: 0 1	Inteiro	Flag para identificar se é uma natureza de defesa civil ou não (0 - Não, 1 - Sim)

NATUREZA1_DESCRIC AO	Variável:	Nominal	Descrição da primeira natureza cadastrada na ocorrência
	Abalo Sísmico Abandono de função Abandono de incapaz Abordagem Abuso de incapazes Achado Afogamento Agressão física/verbal AIFU Alagamento Alarmes Aliciamento de menor Ameaça Animais Antecedentes Criminais - Verificação Apoio Apologia de crime ou criminoso Apreensão de material Apropriação indébita Arrastão Ataque de insetos Atentado violento ao pudor Aterro irregular Atitude suspeita Atos obscenos/libidinosos Avaria em Equipamento/Patrimônio (não intencional) Averiguação Averiguação (Defesa Civil) Banho em local impróprio Bueiro aberto/sem tampa Bueiro entupido Calote Calúnia Câmeras de videomonitoramento Charlatanismo Comércio ambulante Concussão Conduta inconveniente Constrangimento ilegal Construção Irregular Contrabando ou descaminho Corrupção ativa Corrupção de menores Corte irregular de árvore Criança perdida/desaparecida Crime ambiental Dano Denúncia de bomba Depósito irregular Desabamento Desacato Desaparecimento Desinteligência Deslizamento		

	<p> Desobediência  Destelhamento  Devolução de coisa achada  Difamação  Discriminação  Disparo de arma  Embriaguez  Encaminhamento  Envenenamento  Enxurrada  Erosão  Escolta  Escrito ou objeto obsceno (panfletos pornográficos)  Estelionato  Estupro  Exploração de menores  Explosão  Extorsão  Extravio, sonegação ou inutilização de livro ou doc.  Falsidade ideológica (Falsa Identidade)  Favorecimento da prostituição  Fingir-se funcionário público  Fiscalizações e Orientações  Fornecimento de bebida alcoólica à menores  Fuga de aluno/interno  Furto  Homicídio  Homofobia  Impedimento ou perturbação de cerimônia funerária  Importunação ofensiva ao pudor  Importunação sexual  Incêndio  Infiltração  Injúria  Inundação/Enchente  Invasão  Jogo de Azar  Lesão Corporal  Manifestação  Material abandonado  Maus tratos às pessoas  Mendigar, por ociosidade ou cupidez  Menor gazeando aula  Menores abordando transeuntes  Moeda Falsa  Óbito  Óbito (Defesa Civil)  Obstrução de via  Omissão de socorro  Órgãos acionados  Orientação  Paciente/usuário alterado </p>		
--	--	--	--

	<ul style="list-style-type: none"> <li>Peculato</li> <li>Perturbação do sossego</li> <li>Pesca em local proibido</li> <li>Poluição visual/ambiental</li> <li>Porte de artefato explosivo</li> <li>Porte Ilegal</li> <li>Pragas Animais</li> <li>Prostituição</li> <li>Proteção ao patrimônio</li> <li>Queda de aeronave</li> <li>Queda de árvore</li> <li>Queda de fios de energia</li> <li>Queda de galho</li> <li>Queda de poste</li> <li>Quedas de objetos ou partes de construções</li> <li>Queima a céu aberto</li> <li>Racismo</li> <li>Receptação</li> <li>Resistência</li> <li>Retirada de invasão</li> <li>Risco de acidente / à vida</li> <li>Risco de acidente/à vida (Defesa Civil)</li> <li>Risco de desabamento / desmoronamento</li> <li>Risco de explosão</li> <li>Risco de queda de árvore</li> <li>Risco de queda de fios de energia</li> <li>Risco de queda de galho</li> <li>Risco de queda de poste</li> <li>Rixa</li> <li>Rompimento de Barragem</li> <li>Ronda</li> <li>Roubo</li> <li>Rufianismo</li> <li>Saturação</li> <li>Sedução</li> <li>Seqüestro e cárcere privado</li> <li>Situação de risco</li> <li>Substância Ilícita</li> <li>Substância Lícita</li> <li>Suicídio</li> <li>Tentativa</li> <li>Trânsito</li> <li>Usar de uniforme, ou distintivo de função pública que não exerce</li> <li>Uso indevido do cartão transporte</li> <li>Uso indevido do telefone público</li> <li>Vadiagem</li> <li>Vazamento ou derramamento de Produto Perigoso ou Infectante</li> <li>Veículo</li> <li>Venda proibida de produtos específicos à menores</li> <li>Verificação</li> <li>Vias de fato</li> <li>Vilipêndio a cadáver</li> </ul>		
--	---	--	--

	Violação de Medida Protetiva Lei Maria da Penha Violação de sepultura/túmulo Violência arbitrária Vistoria		
OCORRENCIA_ANO	Variável Ex: 2009	Inteiro	Ano de cadastro da ocorrência
OCORRENCIA_CODIGO	Variável Ex: 11516	Inteiro	Código da ocorrência
OCORRENCIA_DATA	Variável Ex: 2009-01-23 11:17:00.000 11516	DATE	Data da ocorrência
OCORRENCIA_DIA_SEMANA	Variável: SEGUNDA TERÇA QUARTA QUINTA SEXTA SÁBADO DOMINGO	Nominal	Dia da semana em que a ocorrência foi cadastrada
OCORRENCIA_HORA	Variável: EX: 19:45:00	time	Hora em que a ocorrência foi cadastrada
OCORRENCIA_MES	Variável: EX: 1	Inteiro	Mês em que a ocorrência foi cadastrada
ORIGEM_CHAMADO_D ESCRICAO	Variável: EX:CIOSP (190)	Alfanumérico	Local em que se originou a chamada
REGIONAL_FATO_NOME	Variável: BAIRRO NOVO BOA VISTA BOQUEIRÃO CAJURU CIC MATRIZ PINHEIRINHO PORTÃO REGIÃO METROPOLITANA SANTA FELICIDADE TATUQUARA	Nominal	Regional do fato (local em que ocorreu a ocorrência)

SECRETARIA_SIGLA	Variável: EX:FAS	Nominal	Sigla da secretaria solicitante
------------------	---------------------	---------	------------------------------------

FONTE: A autora (2020)

## APÊNDICE B - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BOA VISTA

(continua)

Bairro	Total de Ocorrências	Eliminado?
ABRANCHES	1.357	NÃO
AHÚ	2	SIM
ATUBA	1.685	NÃO
BACACHERI	4.837	NÃO
BAIRRO ALTO	2.925	NÃO
BAIRRO FICTICIO	1	SIM
BARREIRINHA	2.296	NÃO
BOA VISTA	3.676	NÃO
CABRAL	6	SIM
CACHOEIRA	602	NÃO
CENTRO	1	SIM
HUGO LANGE	1	SIM
JARDIM SOCIAL	1	SIM
JT TAISA	1	SIM
LOTEAMENTO SÃO GERONIMO	1	SIM
MAUA	1	SIM
PILARZINHO	2.882	NÃO
SANTA CANDIDA	4.127	NÃO
SANTA FELICIDADE	1	SIM
SÃO JOÃO	3	SIM
SÃO LOURENÇO	1.314	NÃO
TABOÃO	1.552	NÃO
TARUMÃ	486	NÃO
TINGUI	870	NÃO
VISTA ALEGRE	2	SIM
TOTAL DE ERROS	21	
TOTAL SEM ERROS	28.609	

FONTE: A autora (2020)

**APÊNDICE C - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL BOQUEIRÃO**

<b>Bairro</b>	<b>Total de Ocorrências</b>	<b>Eliminado?</b>
ALTO BOQUEIRÃO	6.629	NÃO
ALTO DA GLORIA	1	SIM
BOQUEIRÃO	11.977	NÃO
CAJURUR	1	SIM
FANNY	2	SIM
HAUER	2.470	NÃO
JARDIM BOTÂNICO	1	SIM
JATDIM DAS AMERICAS	1	SIM
PAROLIN	1	SIM
SÃO MIGUEL	1	SIM
SITIO CERCADO	1	SIM
UBERABA	1	SIM
XAXIM	4.440	NÃO
TOTAL DE ERROS	10	
TOTAL SEM ERROS	25.516	

FONTE: A autora (2020)

**APÊNDICE D - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL CAJURU**

<b>Bairro</b>	<b>Total de Ocorrências</b>	<b>Eliminado?</b>
BOQUEIRÃO	1	SIM
CAJURU	13.165	NÃO
CAMPO COMPRIDO	1	SIM
CAPÃO DA IMBUIA	2.198	NÃO
CRISTO REI	1	SIM
GUABIRORUBA	1.699	NÃO
JARDIM BOTÂNICO	1	SIM
JARDIM DAS AERICAS	1.160	NÃO
JARDIM SOCIAL	2	SIM
SÃO FRANCISCO	2	SIM
TARUMÃ	51	SIM
UBERABA	5.016	NÃO
TOTAL DE ERROS	59	
TOTAL SEM ERROS	23.238	

FONTE: A autora (2020)



## APÊNDICE E - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL CIC

Bairro	Total de Ocorrências	Eliminado?
AUGUSTA	1.165	NÃO
BIGORRILHO	1	SIM
CAMPINA DO SIQUEIRA	1	SIM
CAMPO COMPRIDO	37	SIM
CAPÃO RASO	1	SIM
CENTRO CIVICO	1	SIM
CIC	2	SIM
CIDADE INDUSTRIAL	17.733	NÃO
FAZENDINHA	4	SIM
NOVO MUNDO	2	SIM
RIVIRA	12	NÃO
SANTA FELICIDADE	2	SIM
SANTA QUITÉRIA	1	SIM
SÃO MIGUEL	572	NÃO
TATUQUARA	1	SIM
TOTAL DE ERROS	51	
TOTAL SEM ERROS	19.484	

FONTE: A autora (2020)

## APÊNDICE F - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL MATRIZ

Bairro	Total de Ocorrências	Eliminado?
ÁGUA VERDE	10	SIM
AHÚ	589	NÃO
ALTO BOQUEIRÃO	1	SIM
ALTO DA GLORIA	1.674	NÃO
ALTO DA RUA XV	1.790	NÃO
BAIRRO FICTICIO	11	SIM
BARREIRINHA	1	SIM
BATEL	1.980	NÃO
BIGORRILHO	2.764	NÃO
BOM RETIRO	525	NÃO
BOQUEIRÃO	1	SIM
CABRAL	2.026	NÃO
CAJURU	29	SIM

CAMPINA DO SIQUEIRA	2	SIM
CAPÃO DA IMBUIA	1	SIM
CASCATINHA	1	SIM
CENTRO	44.160	NÃO
CENTRO CIVICO	2.567	NÃO
COLONIA FARIA	1	SIM
CRISTO REI	998	NÃO
HUGO LANGE	312	NÃO
JARDIM BOTÂNICO	4.247	NÃO
JARDIM SANTA MONICA	1	SIM
JARDIM SOCIAL	372	NÃO
JUVEVÊ	593	NÃO
LINDÓIA	1	SIM
MERCÊS	1.955	NÃO
PAROLIN	5	SIM
PINHEIRINHO	1	SIM
PRADO VELHO	1.569	NÃO
REBOUÇAS	4.325	NÃO
SANTA QUIERIA	1	SIM
SANTO INÁCIO	1	SIM
SÃO FRANCISCO	8.592	NÃO
VISTA ALEGRE	2	SIM
XAXIM	1	SIM
TOTAL DE ERROS	71	
TOTAL SEM ERROS	81.038	

FONTE: A autora (2020)

## APÊNDICE G - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL PORTÃO

Bairro	Total de Ocorrências	Eliminado?
ÁGUA VERDE	4.871	NÃO
BACACHERI	1	SIM
BATEL	2	SIM
BOA VISTA	6	SIM
CAMPINA DO SIQUEIRA	4	SIM
CAMPO COMPRIDO	120	SIM
CAPÃO RASO	1	SIM
CENTRO	3	SIM

CIDADE INDUSTRIAL	5	SIM
FANNY	530	NÃO
FAZENDINHA	5.364	NÃO
GUABIROTUBA	1	SIM
GUAIRA	1.721	NÃO
LINDÓIA	484	NÃO
MERCÊS	1	SIM
NOVO MUNDO	3.408	NÃO
PAROLIN	1.892	NÃO
PORTÃO	5.769	NÃO
REBOUÇAS	3	SIM
SANTA CÂNDIDA	2	SIM
SANTA QUITERIA	991	NÃO
SEMINÁRIO	131	SIM
TINGUI	8	SIM
VILA IZABEL	698	NÃO
TOTAL DE ERROS	288	
TOTAL SEM ERROS	25.728	

FONTE: A autora (2020)

## APÊNDICE H - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL PINHEIRINHO

Bairro	Total de Ocorrências	Eliminado?
ABRANCHES	1	SIM
ÁGUA VERDE	1	SIM
BAIRRO FICTICIO	1	SIM
CAMPO DE SANTANA	1.280	NÃO
CAPÃO RASO	5.768	NÃO
CAXIMBA	213	NÃO
CENTRO	2	SIM
CIDADE INDUSTRIAL	9	SIM
FANNY	249	SIM
GUAIRRA	2	SIM
LINDOIA	213	SIM
NOVO MUNDO	1.149	SIM
PAROLIN	1	SIM
PINHEIRINHO	5.243	NÃO
PORTÃO	6	SIM

PRADO VELHO	1	SIM
SITIO CERCADO	4	SIM
TATUQUARA	6.111	NÃO
UMBARA	1	SIM
XAXIM	5	SIM
TOTAL DE ERROS	1.645	
TOTAL SEM ERROS	18.614	

FONTE: A autora (2020)

## APÊNDICE I - ANÁLISE DE INCONSISTÊNCIA DA REGIONAL SANTA FELICIDADE

Bairro	Total de Ocorrências	Eliminado?
BAIRRO FICTICIO	1	SIM
BATEL	1	SIM
BIGORRILHO	16	SIM
BUTITUVINHA	1.012	NÃO
CMPINA DO SIQUEIRA	2.042	NÃO
CAMPO COMPRIDO	2.713	NÃO
CASCATINHA	218	NÃO
CIDADE INDUSTRIAL	6	SIM
LAMENHA PEQUENA	47	NÃO
MERCÊS	9	SIM
MOSSUNGUÊ	899	NÃO
ORLEANS	84	NÃO
PILARZINHO	7	SIM
SANTA FELICIDADE	4.174	NÃO
SANTO INACIO	4.777	NÃO
SÃO BRAZ	1.640	NÃO
SÃO JOÃO	1.281	NÃO
SEMINARIO	267	NÃO
TINGUI	2	SIM
VISTA ALEGRE	904	NÃO
TOTAL DE ERROS	42	
TOTAL SEM ERROS	20.788	

FONTE: A autora (2020)

## APÊNDICE J- ANÁLISE DE INCONSISTÊNCIA DA REGIÃO METROPOLITANA

Bairro	Quantidade de Dados	Eliminado?
ABRANCHES	1	SIM
ÁGUA VERDE	2	SIM
ÁGUAS BELAS	4	NÃO
ALTO BOQUEIRÃO	2	SIM
ATUBA	1	SIM
BACACHERI	1	SIM
BAIRRO ALTO	1	SIM
BATEL	1	SIM
BOA VISTA	2	SIM
BOQUEIRÃO	3	SIM
BORDA DO CAMPO	3	NÃO
BRAGA	2	NÃO
CAJURU	6	SIM
CAMPO DE SÃO BENEDITO	1	NÃO
CAMPO PEQUENO	3	NÃO
CANGUIRI	1	NÃO
CAPÃO RASO	1	SIM
CENTRO	12	SIM
CIDADE INDUSTRIAL	1	SIM
CIDADE JARDIM	2	NÃO
COLOMBO	1	NÃO
ESTADOS	1	NÃO
FERRARIA	2	NÃO
FORMOSO	1	NÃO
GRALHA AZUL	1	NÃO
IGUAÇÚ 01	1	NÃO
IGUAÇÚ 1	1	NÃO
JARDIM BOA VISTA	1	NÃO
JARDIM BOM PASTOR	1	NÃO
JARDIM COLONIAL	1	NÃO
JARDIM INDUSTRIAL	1	NÃO
JARDIM LOANDA	2	NÃO
JARDIM WEISSOPOLIS	1	NÃO

JD EUROPA	1	NÃO
JD IPE	1	NÃO
JD SUISSA	1	NÃO
LAMENHA PEQUENA	1	NÃO
LOT. MARINONO	1	NÃO
MARACANÃ	1	NÃO
MARIA ANTONIETA	1	NÃO
MENINO DEUS	1	NÃO
NIBTE REY	1	NÃO
NAÇÕES	1	NÃO
NÃO ENCONTRADO	1	NÃO
NÃO INORMADO	3	NÃO
NF	1	NÃO
NI	2	NÃO
NOVO MUNDO	1	SIM
PINHAIS	2	NÃO
PLANTA DEODORO	1	NÃO
PLANTA MEIRELES	1	NÃO
PRADO VELHO	1	SIM
QUATRO BARRAS	1	NÃO
RIO VERDE	1	NÃO
RIVEIRA	1	NÃO
ROÇA NEGRA	1	NÃO
SANTA FELICIDADE	2	SIM
SANTA QUIERIA	1	NÃO
SANTA TEREZINHA	2	NÃO
SANTO ANTONIO	1	NÃO
SÃO CRISTOVÃO	1	NÃO
SÃO GERONIMO	1	NÃO
SÃO JORGE	1	NÃO
SÃO JOSE	1	NÃO
SÃO JOSE OS PINHAIS	1	NÃO
SÃO JUDAS TADEU	1	NÃO
SÃO MIGUEL	1	NÃO
SÃO PEDRO	1	NÃO
SÃO THOMAS	2	NÃO
SEM DADOS	1	NÃO
SITIO DAS PALMEIRAS	1	NÃO
TANGUA	1	NÃO

TATUQUARA	1	SIM
THOMAS COELHO	1	NÃO
UBERABA	1	SIM
VENEZA	1	NÃO
VILABANCARIA	1	NÃO
VILA FORMOSA	2	NÃO
VIL MARIA ANTONIETA	1	NÃO
VILA PERNETA	1	NÃO
VISTA ALEGRE	1	NÃO
XAXIM	1	SIM
TOTAL DE ERROS	60	
TOTAL SEM ERROS	73	

FONTE: A autora (2020)

### APÊNDICE K - AGRUPAMENTO MANUAL DAS DESCRIÇÕES

Grupo_1 (Natureza)	Grupo_2 (apoio)	Grupo_3 (Violência)	Grupo_4 (Dano)	Grupo_5 (Outros)
Abalo_sísmico	Apoio	Abordagem	Acidente_Viatura	Achado
Alagamento	Averiguação	Abuso_de_incapazes	Apropriação_indebita	Alarmes
Animais	Averiguação_(Defesa_Civil)	Agressão_física/verbal	Avaria_em_Equipamento/Patrimônio_(0_intencional)	Banho_em_local_impróprio
Ataque_de_insetos	Câmeras_de_videomonitoramento	Aliciamento_de_menor	Calote	Impedimento_ou_perturbação_de_cerimônia_funerária
Aterro_irregular	Encaminhamento	Ameaça	Conduta_inconveniente	Comércio_ambulante
Corte_irregular_de_árvore	Escolta	Arrastão	Constrangimento_ilegal	Contrabando_ou_desca minho
Crime_ambiental	Órgãos_acionados	Atentado_violento_a_o_pudor	Construção_irregular	Corrupção_ativa
Desabamento	Orientação	Atos_obscenos/libidinosos	Dano	Corrupção_de_menores
Deslizamento	Proteção_ao_patrimônio	Concussão	Difamação	Desacato

Enxurrada	Risco_de_acidente/à_vida_(Defesa_Civil)	Disparo_de_arma	Discriminação	Desobediência
Erosão	Risco_de_acidente/_à_vida	Estupro	Estelionato	Embriaguez
Incêndio	Risco_de_desabamento/_desmoronamento	Furto	Extorsão	Envenenamento
Pesca_em_local_proibido	Risco_de_explosão	Homicídio	Bueiro_aberto/s_em_tampa	Favorecimento_da_prostituição
Poluição_visual/ambiental	Risco_de_queda_de_árvore	Homofobia	Bueiro_entupido	Fingir-se_funcionário_público
Pragas_Animais	Risco_de_queda_de_fios_de_energia	Lesão_corporal	Injúria	Fornecimento_de_bebida_alcoólica_à_menores
Queda_de_árvore	Risco_de_queda_de_galho	Maus_tratamentos_à_pessoas	Obstrução_de_via	Escrito_ou_objeto_obsceeno_(panfletos_pornográficos)
Queda_de_galho	Risco_de_queda_de_poste	Peculato	Perturbação_do_sossego	Jogo_de_azar
Queima_a_céu_aberto	Ronda	Racismo	Queda_de_aeronave	Liberação_de_pessoa_presa/apreendida_por_reusa_no_recebimento_pela_DP
Rompimento_de_Barragem	Vistoria	Roubo	Queda_de_fios_de_energia	Manifestação
Vazamento_ou_derramamento_de_Produto_perigoso_ou_infectante	Criança_perdida/desaparecida	Roubo,_furto,_extravio,_recuperação,_apreensão_de_armas_de_fogo.	Queda_de_poste	Material_abandonado
Infiltração	Denúncia_de_bomba	Seqüestro_e_cárcere_privado	Quedas_de_objetos_ou_partes_de_construções	Mendigar,_por_ociosidade_ou_cupidez
Destelhamento	Desaparecimento	Suicídio	Receptação	Paciente/usuário_alterado
Inundação/Enchente	Devolução_de_coisa_achada	Tentativa	Omissão_de_socorro	Porte_de_artefato_explosivo
	Exploração_de_menores	Calúnia	Rufianismo	Porte_ilegal
	Fuga_de_aluno/interno	Importunação_sexual	Violação_de_sepultura/túmulo	Prostituição
	Retirada_de_invasão	Importunação_ofensiva_ao_pudor	Violência_arbitrária	Quadrilha_ou_bando



Trânsito	Atitude_suspeita	Explosão	Depósito_irregular
Fiscalizações_e_Orientações		Abandono_de_função	Rixa
Fornecimento_de_Lona		Abandono_de_incapaz	Saturação
Situação_de_risco		Invasão	Sedução
Verificação		Moeda_falsa	Substância_ilícita
Menor_gazeando_aula		Vilipêndio_a_cadáver	Substância_lícita
Menores_abordando_transeuntes		Óbito	Usar_de_uniforme,_ou_distintivo_de_função_pública_que_0_exerce
Venda_proibida_de_produtos_específicos_à_menores		Óbito_(Defesa_Civil)	Uso_indevido_do_cartão_transporte
Antecedentes_criminais_Verificação		Extravio,_sonegação_ou_inutilização_de_livro_ou_doc.	Uso_indevido_do_telefone_público
Apologia_de_crime_ou_criminoso			Vadiagem
			Veículo
			Vias_de_fato
			Resistência
			Desinteligência
			AIFU
			Charlatanismo
			Afogamento
			Falsidade_ideológica_(Falsa_Identidade)
			Falsificação_de_documento_publico
			Violação_de_Medida_protetiva_Lei_Maria_da_Penha

FONTE: A autora (2020)

## APÊNDICE L - AGRUPAMENTO DO K-MEANS DAS DESCRIÇÕES

Grupo_1	Grupo_2	Grupo_3	Grupo_4	Grupo_5
Abalo_sísmico	Achado	Alagamento	Abordagem	Agressão física/verbal
Abandono_de_função	AIFU	Atos_obscenos/libidinosos	Apoio	Alarmes
Abandono_de_incipaz	Arrastão	Averiguação	Dano	Ameaça
Abuso_de_incapazes	Banho_em_local_impróprio	Desacato	Perturbação_do_sossego	Animais
Acidente_Viatura	Conduta_inconveniente	Desinteligência	Roubo	Atitude_suspeita
Afogamento	Criança_perdida/desaparecida	Destelhamento	Substância_ilícita	Furto
Aliciamento_de_menor	Crime_ambiental	Encaminhamento		Invasão
Antecedentes_criminais_-_Verificação	Depósito_irregular	Fornecimento_de_Lona		Trânsito
Apologia_de_crime_ou_criminoso	Desabamento	Orientação		
Apropriação_indébita	Desaparecimento	Paciente/usuário_alterado		
Ataque_de_insetos	Desobediência	Queima_a_céu_aberto		
Atentado_violento_ao_pudor	Disparo_de_arma	Risco_de_acidente/_à_vida		
Aterro_irregular	Embriaguez	Ronda		
Avaria_em_Equipamento/Patrimônio_(0_intencional)	Erosão	Saturação		
Averiguação_(Defesa_Civil)	Escolta	Tentativa		
Bueiro_aberto/se_m_tampa	Importunação sexual	Veículo		
Bueiro_entupido	Incêndio	Vias_de_fato		
Calote	Lesão_corporal	Violação_de_Medida_protetiva_Lei_Maria_da_Penha		
Calúnia	Manifestação	Vistoria		

Câmeras_de_vide omonitoramento	Pesca_em_local_p roibido
Charlatanismo	Poluição_visual/a mbiental
Comércio_ambula nte	Porte_ilegal
Concussão	Queda_de_árvore
Constrangimento_ ilegal	Risco_de_desaba mento/_/desmoro namento
Construção_irregu lar	Situação_de_risco
Contrabando_ou_ descaminho	Substância_lícita
Corrupção_ativa	Uso_indevido_do_ cartão_transporte
Corrupção_de_me nores	Vazamento_ou_de rramamento_de_P roduto_perigoso_ ou_infectante
Corte_irregular_d e_árvore	
Denúncia_de_bom ba	
Deslizamento	
Devolução_de_coi sa_achada	
Difamação	
Discriminação	
Envenenamento	
Enxurrada	
Escrito_ou_objeto _obsceno_(panflet os_pornográficos)	
Estelionato	
Estupro	
Exploração_de_m enores	
Explosão	
Extorsão	

Extravio, _sonegação_ou_in utilização_de_livro _ou_doc.
Falsidade_ideológi ca_(Falsa_Identida de)
Falsificação_de_d ocumento_publico
Favorecimento_da _prostituição
Fingir- se_funcionário_pú blico
Fiscalizações_e_Or ientações
Fornecimento_de _bebida_alcoólica _à_menores
Fuga_de_aluno/in terno
Homicídio
Homofobia
Impedimento_ou_ perturbação_de_c erimônia_funerári a
Importunação_ofe nsiva_ao_pudor
Infiltração
Injúria
Inundação/Enchen te
Jogo_de_azar
Liberação_de_pes soa_presa/apreen dida_por_recusa_ no_recebimento_ pela_DP
Material_abandon ado
Maus_tratos_à_pe ssoas
Mendigar, _por_ociosidade_ ou_cupidez

Menor_gazeando_aula
Menores_abordando_transeuntes
Moeda_falsa
Óbito
Óbito_(Defesa_Civil)
Obstrução_de_via
Omissão_de_socorro
Órgãos_acionados
Peculato
Porte_de_artefato_explosivo
Pragas_Animais
Prostituição
Proteção_ao_patrimônio
Quadrilha_ou_bando
Queda_de_aeronave
Queda_de_fios_de_energia
Queda_de_galho
Queda_de_poste
Quedas_de_objetos_ou_partes_de_construções
Racismo
Receptação
Resistência
Retirada_de_invasão
Risco_de_acidente_à_vida_(Defesa_Civil)
Risco_de_explosão
Risco_de_queda_de_árvore
Risco_de_queda_de_fios_de_energia

Risco_de_queda_d e_galho
Risco_de_queda_d e_poste
Rixa
Rompimento_de_ Barragem
Roubo,_furto, _extravio, _recuperação, _apreensão_de_ar mas_de_fogo.
Rufianismo
Sedução
Seqüestro_e_cárc ere_privado
Suicídio
Usar_de_uniforme , _ou_distintivo_de _função_pública_ que_O_exerce
Uso_indevido_do_ telefone_público
Vadiagem
Venda_proibida_d e_produtos_espec íficos_à_menores
Verificação
Vilipêndio_a_cadá ver
Violação_de_sepul tura/túmulo
Violência_arbitrári a

FONTE: A autora (2020)

## APÊNDICE M - RELAÇÃO DE ORIGEM DAS OCORRÊNCIAS POR ANOS

Origem	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
153	7.135	7.944	6.498	7.934	11.280	11.502	12.428	8.881	8.743	9.235	9.706	101.286
IMEDIATA	3.338	3.163	3.867	4.186	4.197	3.473	3.351	4.284	6.280	7.194	14.945	58.278
CIOSP (190)	4.718	2.769	1.482	1.601	1.631	2.007	2.082	1.862	1.639	1.371	827	21.989
AOS GMS	1.824	1.670	2.172	1.596	2.019	2.450	1.983	1.587	1.288	1.230	2.092	19.911
156	2.365	1.893	1.647	1.208	1.792	1.733	924	302	119	612	884	13.479
À VIATURA	1.179	805	955	866	1.123	1.267	1.859	1.126	1.069	982	1.600	12.831
FONE	887	809	740	846	899	1.078	1.103	972	803	1.009	570	9.716
G5	355	373	529	382	712	867	791	755	743	516	330	6.353
SIGA	2.308	1.786	1.503	210	5	11	13	2	2	69	15	5.924
OFÍCIO	106	528	1.341	361	202	261	257	174	160	140	73	3.603
CCO URBS	363	487	477	331	238	261	402	167	54	223	338	3.341
SAMU	5	11	8	7	33	177	145	357	567	564	707	2.581
199	138	137	128	67	32	12	17	7	7	43	124	712
CCO RUAS DA CIDADANIA	22	25	18	85	205	125	79	34	23	48	27	691
CCO BARIGUI	79	48	15	29	9	6	5		1			192
CCO MUNICIPAL URBS							10	7		130	23	170
EMAIL			8	2	13	12	22	20	22	14	14	127
NRGM											124	124
REDE SOCIAL							13	14		6	18	51
CASA DA MULHER BRASILEIRA								8	9	11	11	39
TOTEM RUA SÃO FRANSICO								27	3	3		33
CCO SÍTIO CERCADO					10	7	7	1	1	1		27
VEPPER											20	20
CCO PUC		6		2	3	1	3		1		1	17
CENTRAL DE FLAGRANTES											2	2
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

## APÊNDICE N- RELAÇÃO DE OCORRÊNCIAS POR SECRETARIA EM ANOS

Secretaria	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
SMDT	8.460	9.341	12.898	9.758	12.262	14.011	11.288	10.715	11.541	12.819	24.625	137.718
SMMA	7.011	5.966	2.354	4.057	5.821	5.321	5.527	3.347	3.336	3.127	2.839	48.706
URBS	3.094	2.171	2.153	2.654	2.944	2.990	4.486	2.835	2.554	2.362	1.458	29.701
SMS	1.688	998	1.152	1.287	1.458	1.163	1.713	1.491	1.471	1.591	1.204	15.216
SME	1.107	1.136	1.098	789	930	957	1.303	1.212	1.241	1.229	848	11.850
FAS	2.818	2.288	1.350	662	504	271	416	336	458	618	487	10.208
SMU	315	178	63	201	177	212	365	244	375	299	217	2.646
SETRAN				25	47	50	53	52	21	792	167	1.207
SMSAN	78	97	106	101	63	72	75	81	81	71	47	872
SMEM								80	189	202	193	664
COHAB	89	139	94	76	49	23	48	30	10	15	12	585
SMELJ	43	36	31	30	42	46	72	54	55	59	69	537
SMCS	22	25	31	22	37	66	38	24	62	56	122	505
SMOP	38	31	23	16	25	15	36	22	23	34	14	277
SGM	4		3	2	2	1	16	5	57	43	70	203
SMAP	6	3	5	10	4	21	21	27	30	33	26	186
FCC	21	16	11	9	13	9	11	8	11	14	12	135
SMF	5	8	1	2	7	5	2	4	7	9	21	71
SMCR	4	8	11	5	1	1	1	3	1	3	5	43
SAM	4	1			8	5	3	4	1	5		31
SMRH	1	2		1	3	1	2		4	7	4	25
SMSA	4	2	1	1	1	2		2			1	14
SEDPD					1		4	2	1	3	1	12
SMTE	1	2	2			2	3	1		1		12
CSA	1	3		2		1	1		2		1	11
SMTU	5	1		1		1				2		10
SMPC	1	2		1	1	1	1			1		8
SERIN						1		1		4		6
SICT							4	1			1	6
SMAM				1	1			1	1	1	1	6
IMAP					1			1		1	2	5
SECOPA							2	2				4
SEPLAN								1	1		2	4
SEA						2	1					3
SEPHA					1						2	3
SERIC	2								1			3
SEPE							1	1				2
SERCOM			1				1					2
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>



## APÊNDICE O - RELAÇÃO DE DESCRIÇÕES POR ANOS

Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
Apoio	2.397	2.641	2.656	2.708	3.163	3.769	3.567	3.629	4.232	4.592	6.379	39.733
Dano	3.093	2.699	3.116	3.448	4.680	3.966	3.600	2.392	2.146	1.990	1.875	33.005
Abordagem	433	678	865	747	832	1.169	1.467	1.583	2.161	2.354	6.927	19.216
Perturbação do sossego	5.481	4.637	983	973	1.092	1.561	1.327	881	785	668	602	18.990
Substância Ilícita	652	1.064	1.393	2.038	2.427	1.719	1.589	1.002	1.179	1.271	2.012	16.346
Roubo	531	537	699	862	1.143	2.222	2.655	2.375	1.965	1.497	1.302	15.788
Animais	2.198	1.988	1.823	1.399	2.129	1.721	1.031	331	114	97	162	12.993
Invasão	924	776	750	885	1.108	1.155	1.458	915	1.151	1.088	1.061	11.271
Alarmes	2.449	1.927	1.616	410	627	840	602	467	394	402	171	9.905
Atitude suspeita	273	384	614	699	1.148	991	1.120	978	1.096	1.014	1.355	9.672
Trânsito	614	673	758	692	720	597	558	442	472	1.549	1.258	8.333
Ameaça	242	246	311	452	477	743	1.214	944	696	635	590	6.550
Furto	372	370	409	410	529	595	566	484	595	757	814	5.901
Agressão física/verbal	272	295	316	378	378	472	563	456	458	475	624	4.687
Vias de fato	297	293	297	325	390	470	448	272	238	324	356	3.710
Veículo	60	126	206	170	220	263	386	358	224	301	342	2.656
Queima a céu aberto	158	192	185	169	235	207	242	272	252	341	339	2.592
Tentativa	97	149	134	166	212	236	255	240	260	248	276	2.273
Alagamento	526	314	572	167	93	134	132	53	61	142	64	2.258
Desinteligência	81	63	85	126	211	161	197	182	250	284	484	2.124
Orientação	120	100	151	131	109	132	154	218	150	197	595	2.057
Violação de Medida Protetiva Lei Maria da Penha								172	512	538	602	1.824
Vistoria	5	68	322	144	62	69	72	100	143	260	379	1.624
Encaminhamento	206	144	96	74	108	93	87	91	152	146	359	1.556
Atos obscenos/libidinosos	121	156	134	152	178	153	139	106	121	124	138	1.522
Destelhamento	34	58	127	299	357	60	91	81	20	33	215	1.375
Paciente/usuário alterado	44	68	88	66	55	46	94	187	172	284	253	1.357
Ronda	30	76	120	86	153	113	158	102	169	117	109	1.233
Averiguação	906	324										1.230
Risco de acidente / à vida	105	56	99	65	166	161	152	83	55	102	147	1.191
Desacato	88	112	118	122	114	118	109	77	71	104	105	1.138
Saturação	8	7	1.013									1.028
Fornecimento de Lona											1.017	1.017
Pesca em local proibido	108	39	72	93	30	71	100	81	101	80	50	825
Disparo de arma						162	175	141	100	139	95	812
Lesão Corporal		36	59	74	83	91	116	92	76	79	53	759
Queda de árvore	106	83	86	76	74	32	28	19	29	65	65	663
AIFU	622	40										662
Achado	26	40	48	63	46	43	55	39	45	63	134	602
Porte Ilegal	38	47	49	53	70	53	49	49	53	48	37	546

Uso indevido do cartão transporte	54	27	35	28	42	30	49	27	50	51	66	459
Banho em local impróprio	65	54	37	71	25	49	34	17	28	30	42	452
Risco de desabamento / desmoronamento	66	64	83	61	43	24	24	17	9	26	20	437
Conduta inconveniente	16	25	47	95	56	18	15	21	37	38	58	426
Desabamento	57	46	75	56	34	24	28	12	6	37	19	394
Incêndio	43	36	42	36	39	37	39	23	28	25	31	379
Depósito irregular	18	23	33	34	46	43	40	24	35	38	37	371
Importunação sexual	10	10	15	24	23	27	64	47	53	43	54	370
Criança perdida/desaparecida	28	16	29	27	47	41	30	26	23	26	49	342
Substância Lícita	50	66	43	48	46	23	9	11	16	3	11	326
Desaparecimento	30	25	25	22	27	23	27	24	29	33	35	300
Crime ambiental	16	26	25	19	33	18	26	27	31	27	40	288
Escolta	13	7	8	11	24	24	53	36	46	45	15	282
Manifestação	16	8	25	21	34	35	14	12	35	44	30	274
Erosão	72	46	35	28	13	12	16	21	3	11	6	263
Situação de risco	23	20	19	11	32	38	14	7	14	22	21	221
Arrastão	2	7	3	6	6	10	48	37	29	52	12	212
Embriaguez	25	18	16	25	21	30	10	14	15	12	19	205
Poluição visual/ambiental	50	19	21	14	18	12	11	3	6	15	24	193
Desobediência	9	9	8	15	9	30	17	11	19	30	33	190
Vazamento ou derramamento de Produto Perigoso ou Infectante	30	22	22	15	13	5	18	9	8	24	16	182
Avaria em Equipamento/Patrimônio (não intencional)	4	12	10	19	13	7	15	15	17	17	20	149
Corte irregular de árvore	14	12	21	11	16	12	10	7	14	13	12	142
Óbito	16	19	17	10	3	17	13	11	7	12	13	138
Abandono de incapaz	14	8	13	11	11	12	15	5	14	15	16	134
Retirada de invasão	18	34	9	5	5	7	17	12	8	8	11	134
Averiguação (Defesa Civil)	12	21	15	14	11	8	5	6	10	20	9	131
Rixa	12	11	16	14	5	7	12	5	12	6	10	110
Quedas de objetos ou partes de construções	18	15	11	7	9	11	8	7	5	7	11	109
Fuga de aluno/interno	6	6	6	6	11	6	15	10	8	9	25	108
Estelionato	5	10	13	16	8	11	8	5	8	7	12	103
Obstrução de via	2	7	11	6	8	5	10	1	10	24	19	103
Homicídio	9	14	9	12	10	7	10	4	8	7	5	95
Construção Irregular	11	10	5	2	12	3	5	10	10	11	12	91
Risco de acidente/à vida (Defesa Civil)	12	5	9	4	9	2	5	8	5	15	14	88
Importunação ofensiva ao pudor	4	4	4	1	4	4	12	7	15	15	17	87
Deslizamento	15	13	14	2	5	7	4	3	3	10	6	82
Estupro	3	5	3	7	7	9	12	14	10	3	6	79
Verificação	3	7	10	4	6	6	5	5	5	9	12	72



Impedimento ou perturbação de cerimônia funerária		1			3	1	1	4	1	1	6	18
Racismo			4	1	1	1	4	1	2	2	1	17
Bueiro aberto/sem tampa	3	4		1	1	3		1	1		2	16
Risco de explosão	1	1	3	3		1	1	1		3	2	16
Calúnia	3			2		1	1	1	1	1	3	13
Usar de uniforme, ou distintivo de função pública que não exerce	1				1	2	4	1	2	2		13
Exploração de menores		1	2	2	1	1	2	1		1	1	12
Bueiro entupido	4			2	1	2		1		1		11
Violação de sepultura/túmulo	1	2	1				1		1	1	4	11
Contrabando ou descaminho	1		1		1		1		5	1		10
Favorecimento da prostituição	4	3	1	2								10
Acidente Viatura											9	9
Risco de queda de poste		3		2	1	1		1		1		9
Aterro irregular	1			1	2		1	1			1	7
Órgãos acionados		2		1		1			1	1	1	7
Risco de queda de fios de energia	2		1		1	1			1	1		7
Corrupção de menores		1		1	1			1		1	1	6
Homofobia						1				2	3	6
Inundação/Enchente			1	1		1	1			2		6
Pragas Animais	1			1	1				2		1	6
Risco de queda de galho	1		2	2	1							6
Vilipêndio a cadáver							1		4	1		6
Uso indevido do telefone público		1	1						1	1	1	5
Venda proibida de produtos específicos à menores					2	1	1		1			5
Abandono de função			1		1						2	4
Câmeras de videomonitoramento			2		1					1		4
Enxurrada			1					1	1		1	4
Fingir-se funcionário público				1		1		1		1		4
Fornecimento de bebida alcoólica à menores	1	2							1			4
Peculato	3	1										4
Charlatanismo		2		1								3
Difamação			2				1					3
Discriminação						1			1	1		3
Omissão de socorro										2	1	3
Queda de aeronave	1		1			1						3
Sedução			1		1						1	3
Violência arbitrária			1		2							3
Abalo Sísmico					2							2
Apologia de crime ou criminoso	1					1						2

Concussão		1								1		2
Envenenamento		1					1					2
Corrupção ativa		1										1
Falsificação de documento publico											1	1
Liberação de pessoa presa/apreendida por recusa no recebimento pela DP											1	1
Óbito (Defesa Civil)						1						1
Prostituição	1											1
Quadrilha ou bando											1	1
Rompimento de Barragem									1			1
Rufianismo					1							1
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

## APÊNDICE P - RELAÇÃO DE BAIRROS POR ANOS

Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
CENTRO	2.633	2.595	3.833	3.373	3.839	3.921	3.874	2.768	3.507	4.024	9.793	44.160
CIDADE INDUSTRIAL	2.068	1.859	1.527	1.133	1.368	1.746	1.865	1.358	1.497	1.724	1.590	17.735
SÍTIO CERCADO	1.600	1.546	1.488	1.202	1.322	1.468	1.469	1.487	1.430	1.452	1.597	16.061
CAJURU	1.396	1.213	962	838	1.125	1.176	1.239	1.153	1.151	1.288	1.624	13.165
BOQUEIRÃO	1.246	939	1.096	936	1.005	1.008	1.027	1.115	1.211	1.114	1.280	11.977
SÃO FRANCISCO	421	618	715	668	968	825	604	364	805	821	1.783	8.592
ALTO BOQUEIRÃO	620	465	465	509	594	624	706	643	687	688	628	6.629
TATUQUARA	683	557	430	358	632	490	525	583	527	584	742	6.111
PORTÃO	563	516	464	465	600	616	635	487	403	485	535	5.769
CAPÃO RASO	481	472	486	459	531	648	667	571	437	498	518	5.768
FAZENDINHA	553	459	408	504	492	603	563	441	423	407	511	5.364
PINHEIRINHO	633	573	453	390	496	581	592	406	327	386	406	5.243
UBERABA	735	594	417	374	388	408	362	406	392	433	507	5.016
ÁGUA VERDE	356	337	360	394	490	458	583	454	475	542	422	4.871
BACACHERI	339	316	371	511	503	467	491	426	453	468	492	4.837
SANTO INÁCIO	465	446	332	498	492	560	472	344	341	341	486	4.777
XAXIM	465	404	440	289	348	617	513	374	319	353	318	4.440
REBOUÇAS	297	370	313	332	479	435	494	390	359	332	524	4.325
JARDIM BOTÂNICO	245	211	243	238	355	369	371	339	428	549	899	4.247
SANTA FELICIDADE	521	450	371	258	425	367	414	309	287	359	413	4.174
SANTA CÂNDIDA	517	475	349	251	367	307	391	373	366	268	463	4.127
BOA VISTA	452	430	281	242	348	323	333	289	266	242	470	3.676
NOVO MUNDO	681	536	411	362	482	479	445				12	3.408
BAIRRO ALTO	439	438	257	216	287	209	254	185	206	203	231	2.925
PILARZINHO	353	301	269	192	251	223	277	247	229	232	308	2.882
BIGORRILHO	161	140	134	185	257	330	319	222	321	316	379	2.764
CAMPO COMPRIDO	384	329	224	249	309	290	269	195	188	189	87	2.713

CENTRO CÍVICO	223	206	210	173	228	227	197	158	209	285	451	2.567
HAUER	244	181	195	153	222	237	240	283	227	210	278	2.470
BARREIRINHA	303	251	193	188	233	255	234	152	171	143	173	2.296
CAPÃO DA IMBUIA	237	207	177	234	154	181	236	203	192	188	189	2.198
CAMPINA DO SIQUEIRA	164	164	131	143	205	245	214	182	206	214	174	2.042
CABRAL	300	198	170	163	177	197	218	145	114	169	175	2.026
BATEL	104	91	63	99	206	179	205	162	175	352	344	1.980
MERCÊS	188	161	145	161	249	239	204	159	139	140	170	1.955
PAROLIN	173	174	184	111	101	160	217	269	191	159	153	1.892
ALTO DA RUA XV	94	88	67	115	160	238	215	195	227	199	192	1.790
GUÁIRA	175	177	162	102	164	138	119	187	177	217	103	1.721
GUABIROTUBA	78	128	60	171	192	314	207	145	154	121	129	1.699
ATUBA	174	145	132	120	217	163	132	136	121	203	143	1.686
ALTO DA GLÓRIA	136	114	95	127	169	154	278	232	150	90	129	1.674
SÃO BRAZ	186	200	170	123	169	174	127	110	75	88	218	1.640
PRADO VELHO	125	94	114	146	170	183	166	175	137	146	114	1.570
TABOÃO	82	74	154	198	186	164	134	116	156	149	139	1.552
ABRANCHES	123	117	85	127	115	153	149	116	107	130	135	1.357
SÃO LOURENÇO	99	110	90	112	197	130	165	106	116	106	83	1.314
SÃO JOÃO	76	104	98	71	99	126	178	150	106	148	125	1.281
CAMPO DE SANTANA	214	206	111	117	224	186	222					1.280
UMBARÁ	164	158	122	101	109	76	107	97	93	115	118	1.260
AUGUSTA	75	100	76	82	113	92	91	130	137	148	121	1.165
JARDIM DAS AMÉRICAS	125	88	75	122	150	123	97	75	69	109	127	1.160
GANCHINHO	141	92	45	50	88	96	126	114	91	100	141	1.084
BUTIATUVINH A	94	94	127	61	88	132	60	74	92	97	93	1.012
CRISTO REI	64	73	70	105	132	94	108	92	57	73	130	998
SANTA QUITÉRIA	123	97	84	55	112	114	83	59	50	106	109	992
VISTA ALEGRE	109	91	98	90	89	72	83	55	61	82	75	905
MOSSUNGUÊ	75	62	64	111	116	102	86	72	70	69	72	899
TINGUI	92	68	73	38	44	55	55	70	77	130	168	870
ORLEANS	141	59	56	34	28	37	52	45	58	80	224	814
VILA IZABEL	68	48	39	42	101	65	85	75	69	50	56	698
CACHOEIRA	91	74	76	34	65	45	44	29	43	62	39	602
JUVEVÊ	81	68	54	46	50	44	53	35	46	60	56	593
AHÚ	48	64	49	65	63	46	54	42	37	77	44	589
SÃO MIGUEL	32	45	32	28	51	53	53	51	81	86	61	573
FANNY	83	65	64	50	75	95	95				3	530
BOM RETIRO	64	66	44	52	64	41	47	25	45	40	37	525
TARUMÃ	48	41	45	52	54	55	42	29	70	39	11	486
LINDÓIA	78	59	56	29	69	73	92				28	484
JARDIM SOCIAL	29	26	34	23	43	28	31	22	28	63	45	372
HUGO LANGE	22	23	18	16	21	36	55	49	31	22	19	312
SEMINÁRIO	53	62	37	19	48	23	23				2	267
CASCATINHA	28	21	23	9	13	23	24	25	21	12	19	218
CAXIMBA	80	30	20	14	23	23	23					213
LAMENHA PEQUENA	6	1		2	2	15	5	4	2	9	2	48



SANTO ANTONIO								1				1
SÃO CRISTOVÃO									1			1
SAO GERONIMO											1	1
SÃO JORGE									1			1
SÃO JOSE								1				1
SAO JOSE DOS PINHAIS									1			1
SÃO JUDAS TADEU								1				1
SÃO PEDRO									1			1
SÃO THOMAS										1		1
SÃO THOMAZ										1		1
SEM DADOS									1			1
SITIO DAS PALMEIRAS										1		1
TANGUA							1					1
THOMAS COELHO										1		1
VENEZA							1					1
VILA BANCÁRIA										1		1
VILA MARIA ANTONIETA								1				1
VILA PERNETA									1			1
Total Geral	24.822	22.454	21.388	19.713	24.403	25.250	25.494	20.587	21.534	23.401	32.451	261.497

## APÊNDICE Q - RELAÇÃO DA DESCRIÇÃO COM SUBCATEGORIA

Rótulos de Linha	Abordagem	Alarques	Animais	Apoio	Atitude suspeita	Dano	Invasão	Perturbação do sossego	Roubo	Substância Ilícita	Total Geral
Abordagem	19.216										19.216
Perturbação do sossego								18.990			18.990
Uso de substância ilícita										11.684	11.684
Apoio ao cidadão - PRESTAÇÃO DE SOCORRO/SALVAMENTO				10.915							10.915
Atitude suspeita					9.672						9.672
Pichação						9.149					9.149
Disparo de Alarme (violação)		9.088									9.088
Vandalismo						8.885					8.885
Cão solto em via pública			8.769								8.769
Transporte Coletivo									7.452		7.452
Desordem						6.136					6.136
Transeunte									6.008		6.008
Arrombamento						5.830					5.830



Apoio ao SAMU				5.629							5.629
Invasão de equipamento/patrimônio público							5.620				5.620
Apoio à FAS				4.107							4.107
Invasão ao transporte coletivo							4.100				4.100
Apoio ao cidadão - ORIENTAÇÃO				3.373							3.373
Maus tratos a animais			3.073								3.073
Tráfico de drogas										2.960	2.960
Tumulto						2.898					2.898
Apoio ao GM				2.841							2.841
Apoio a outros órgãos				2.512							2.512
Apoio à URBS				1.725							1.725
Porte de substância ilícita										1.702	1.702
Apoio à SMS				1.635							1.635
Apoio à SETRAN				1.268							1.268
Apoio à SME				1.132							1.132
Apoio à SMU				979							979
Invasão de terreno							822				822
Estabelecimento Comercial									729		729
Apoio à PMPR				714							714
Veículo									647		647
Equipamento/patrimônio público									630		630
Apoio ao Conselho Tutelar				540							540
Apoio à COHAB				520							520
Cão sem guia/focinheira			469								469
Apoio à SMMA				453							453
Apoio à Polícia Civil				436							436
Falha na comunicação		376									376
Ocupação irregular							349				349
Apoio ao SIATE				255							255
Socorro e encaminhamento de animais/aves			243								243
Falha no sistema		239									239
Apoio ao Corpo de Bombeiros				210							210
Invasão de propriedade particular							196				196
Invasão de residência							184				184
Uso de bebida alcoólica											184
Residência									163		163
Auto teste não recebido		156									156
Equídeo solto em via pública			155								155
Arma de Fogo									135		135
Cão preso inadequadamente			129								129
Apoio ao Patrimônio				107							107
Grafitagem Autorizada						107					107
Apoio à SMEL				87							87
Apoio à SMOP				79							79

Uso de cola/thinner/similar											76
Omissão de Cautela na Guarda de animais			62								62
Apoio à SMAB				58							58
Apoio à Vigilância Sanitária				53							53
Falta de energia elétrica		46									46
Comércio ilegal de aves/animais			39								39
Criador ilegal de aves/animais			29								29
Porte de cola/thinner/similar											29
Apoio à Polícia Rodoviária Federal				28							28
Apoio à COSEDI				27							27
Dejetos gerados por animais de estimação			25								25
Apoio à SMAD				23							23
Instituição Financeira									20		20
Porte de bebida alcoólica											20
Apoio à Polícia Federal				17							17
Oferecer bebida alcoólica a menor de 18 anos											16
Apoio ao IML				5							5
Instituição Religiosa									4		4
Apoio à SAM				3							3
Apoio ao Horto Municipal				2							2
Vender bebida alcoólica a menor de 18 anos											1
<b>Total Geral</b>	<b>19.216</b>	<b>9.905</b>	<b>12.993</b>	<b>39.733</b>	<b>9.672</b>	<b>33.005</b>	<b>11.271</b>	<b>18.990</b>	<b>15.788</b>	<b>16.346</b>	<b>187.245</b>

## APÊNDICE R – RELAÇÃO SUBCATEGORIAS POR ANOS

Rótulos de Linha	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total Geral
Abordagem	433	678	865	747	832	1.169	1.467	1.583	2.161	2.354	6.927	19.216
Perturbação_do_sossego	5.481	4.637	983	973	1.092	1.561	1.327	881	785	668	602	18.990
Uso_de_substância_ilícita	525	862	1.127	1.620	1.777	1.180	1.036	669	828	807	1.253	11.684
Apoio_ao_cidadão_-_PRESTAÇÃO_DE_SOCORRO/SALVAMENTO	967	835	974	961	1.032	912	926	698	905	1.104	1.601	10.915
Atitude_suspeita	273	384	614	699	1.148	991	1.120	978	1.096	1.014	1.355	9.672
Pichação	685	643	713	1.148	2.132	1.423	1.109	521	226	242	307	9.149
Disparo_de_Alarme_(violação)	2.375	1.895	1.570	391	546	665	446	355	327	361	157	9.088
Vandalismo	791	770	790	874	1.050	1.045	921	703	659	634	648	8.885
Cão_solto_em_via_pública	2.060	1.845	1.679	1.266	1.105	228	388	160	14	8	16	8.769
Transporte_coletivo	189	146	284	356	552	1.252	1.505	1.315	973	680	508	7.760
Transeunte	414	433	411	481	540	819	945	847	920	812	890	7.512
Ameaça	242	246	311	452	477	743	1.214	944	696	635	590	6.550
Desordem	987	617	698	656	701	577	646	319	309	285	341	6.136
Arrombamento	322	406	683	535	501	548	595	581	703	587	369	5.830

Apoio_ao_SAMU	75	113	107	171	275	608	457	764	983	960	1.116	5.629
Invasão_de Equipamento/ patrimônio público	333	357	393	470	565	548	624	491	611	584	644	5.620
Acidente_de trânsito	416	476	538	504	474	423	395	323	338	450	571	4.908
Agressão_física/verbal	272	295	316	378	378	472	563	456	458	475	624	4.687
Apoio_à_FAS	520	497	395	366	266	218	205	189	391	468	592	4.107
Invasão_ao transporte coletivo	441	318	273	295	418	445	608	287	377	351	287	4.100
Vias_de_fato	297	293	297	325	390	470	448	272	238	324	356	3.710
Apoio_ao cidadão - ORIENTAÇÃO	143	135	150	140	231	324	353	318	309	325	945	3.373
Maus_tratos_a_animais	14	11	14	17	840	1.409	545	118	27	29	49	3.073
Tráfico_de drogas	42	96	169	290	463	397	413	229	227	249	385	2.960
Tumulto	308	263	232	235	283	324	305	258	244	239	207	2.898
Apoio_ao_GM	116	127	208	224	281	369	347	337	333	242	257	2.841
Equipamento/patrimônio público	132	116	182	154	194	299	279	282	309	452	352	2.751
Queima_a_céu_aberto	158	192	185	169	235	207	242	272	252	341	339	2.592
Apoio_à_outros_órgãos	44	247	225	102	136	242	207	245	196	231	637	2.512
Alagamento	526	314	572	167	93	134	132	53	61	142	64	2.258
Desinteligência	81	63	85	126	211	161	197	182	250	284	484	2.124
Orientação	120	100	151	131	109	132	154	218	150	197	595	2.057
Violação_de Medida protetiva Lei_Maria_da_Penha								172	512	538	602	1.824
Apoio_à_URBS	134	100	85	189	239	315	161	123	124	113	142	1.725
Porte_de substância ilícita	85	106	97	128	187	142	140	104	124	215	374	1.702
Apoio_à_SMS	64	46	55	73	103	156	231	243	243	187	234	1.635
Vistoria	5	68	322	144	62	69	72	100	143	260	379	1.624
Estabelecimento comercial	69	89	111	116	197	204	224	175	150	119	152	1.606
Estacionamento irregular	23	34	24	22	34	17	23	9	28	888	445	1.547
Atos obscenos/libidinosos	121	156	134	152	178	153	139	106	121	124	138	1.522
Veículo	75	87	88	111	122	178	169	162	161	145	152	1.450
Destelhamento	34	58	127	299	357	60	91	81	20	33	215	1.375
Paciente/usuário alterado	44	68	88	66	55	46	94	187	172	284	253	1.357
Recuperação_de veículo furtado/roubado	25	58	82	84	116	171	250	239	122	85	90	1.322
Apoio_à_SETRAN	40	105	78	81	87	110	66	46	44	353	258	1.268
Ronda	30	76	120	86	153	113	158	102	169	117	109	1.233
Averiguação	906	324										1.230
Desacato	88	112	118	122	114	118	109	77	71	104	105	1.138
Apoio_à_SME	29	24	29	32	73	95	141	161	200	143	205	1.132
Saturação	8	7	1.013									1.028
Fornecimento_de Lona											1.017	1.017
Apoio_à_SMU	75	144	30	96	71	54	70	65	150	121	103	979
Encaminhamento_de foragido (mandado de busca e apreensão)	16	35	31	24	73	66	63	70	135	126	325	964
Tentativa_de roubo	32	48	51	58	83	112	116	133	120	106	104	963
Pesca_em_local proibido	108	39	72	93	30	71	100	81	101	80	50	825
Invasão_de terreno	102	47	48	58	58	86	120	84	94	79	46	822
Veículo aberto/abandonado	25	47	94	66	80	57	72	61	47	84	82	715

Apoio_à_PMPR	33	45	65	46	69	72	84	109	72	52	67	714
Direção_perigosa	57	60	69	48	78	71	51	50	41	71	111	707
Queda_de_árvore	106	83	86	76	74	32	28	19	29	65	65	663
AIFU	622	40										662
Atropelamento	89	65	79	75	74	46	55	31	35	57	47	653
Apoio_ao_Conselho_tutelar	38	33	41	52	48	49	37	55	56	75	56	540
Apoio_à_COHAB	57	100	97	57	31	29	43	43	25	17	21	520
Arma_de_Fogo	24	29	31	47	61	52	83	52	45	45	28	497
Cão_sem_guia/focinheira	71	69	53	53	42	20	28	16	30	30	57	469
Disparo_de_arma_por_GM_com_munição_menos_letal						83	112	58	51	105	60	469
Por_arma_de_fogo		25	49	46	54	61	83	61	31	44	15	469
Uso_indevido_do_cartão_transporte	54	27	35	28	42	30	49	27	50	51	66	459
Apoio_à_SMMA	24	13	11	13	74	55	43	58	49	68	45	453
Banho_em_local_impróprio	65	54	37	71	25	49	34	17	28	30	42	452
Situação_irregular	10	21	30	20	24	35	29	26	30	106	116	447
Surf_em_veículo_coletivo	25	25	26	15	78	85	97	32	12	34	16	445
Tentativa_de_furto	19	36	30	33	45	39	37	31	54	55	64	443
Risco_de_desabamento/_/desmoronamento	66	64	83	61	43	24	24	17	9	26	20	437
Apoio_à_Polícia_Civil	6	20	19	21	29	39	63	68	75	66	30	436
Conduta_inconveniente	16	25	47	95	56	18	15	21	37	38	58	426
Residência	20	29	26	35	41	45	43	59	33	37	27	395
Desabamento	57	46	75	56	34	24	28	12	6	37	19	394
Incêndio	43	36	42	36	39	37	39	23	28	25	31	379
Falha_na_comunicação	34	9	24	8	41	62	77	53	37	21	11	377
Depósito_irregular	18	23	33	34	46	43	40	24	35	38	37	371
Importunação sexual	10	10	15	24	23	27	64	47	53	43	54	370
Ocupação_irregular	30	29	12	23	17	36	55	21	39	37	50	349
Criança_perdida/desaparecida	28	16	29	27	47	41	30	26	23	26	49	342
Pipa_com_cerol	64	11	15	18	45	39	32	42	24	9	33	332
Encaminhamento_hospitalar	112	76	35	28	17	15	13	12	8	6	7	329
Condutor_embriagado	22	22	28	16	33	26	20	23	18	50	60	318
Desaparecimento	30	25	25	22	27	23	27	24	29	33	35	300
Crime_ambiental	16	26	25	19	33	18	26	27	31	27	40	288
Escolta	13	7	8	11	24	24	53	36	46	45	15	282
Manifestação	16	8	25	21	34	35	14	12	35	44	30	274
Erosão	72	46	35	28	13	12	16	21	3	11	6	263
Rabeira_em_veículo_coletivo	6	5	36	17	14	18	10	6	9	52	87	260
Apoio_ao_SIATE	15	33	42	17	26	28	35	25	16	8	10	255
Socorro_e_encaminhamento_de_animais/aves	15	21	36	25	40	22	25	10	18	14	17	243
Falha_no_sistema	8	12	13	10	33	72	48	23	13	9	1	242
Situação_de_risco	23	20	19	11	32	38	14	7	14	22	21	221
Arrastão	2	7	3	6	6	10	48	37	29	52	12	212
Apoio_ao_Corpo_de_Bombeiros			7	31	32	29	37	21	20	11	22	210
Embriaguez	25	18	16	25	21	30	10	14	15	12	19	205

Por_arma_branca		7	8	25	28	21	21	21	32	23	19	205
Disparo_de_arma_por_terceiro						49	31	55	27	18	19	199
Invasão_de_propriedade_particular	9	9	12	21	32	17	27	18	18	19	14	196
Achado_de_objeto	9	9	10	11	19	16	12	6	10	18	73	193
Poluição_visual/ambiental	50	19	21	14	18	12	11	3	6	15	24	193
Desobediência	9	9	8	15	9	30	17	11	19	30	33	190
Invasão_de_residência	9	16	12	18	18	23	24	14	12	18	20	184
Uso_de_bebida_alcoólica	27	45	22	22	25	16	6	3	11	3	4	184
Vazamento_ou_derramamento_de_Produto_perigoso_ou_infectante	30	22	22	15	13	5	18	9	8	24	16	182
Condutor_O_habilitado	5	15	18	26	23	12	11	4	12	32	21	179
Tentativa_de_invasão_transportecoletivo	3	12	5	14	23	10	39	11	30	16	10	173
Verificação_da_Situação_do_veículo							35	32	25	26	54	172
Autoteste_O_recebido	18		7	1	6	37	28	32	14	11	2	156
Equídeo_solto_em_via_pública	13	9	11	10	39	13	12	15	12	9	12	155
Tentativa_de_arrombamento	4	19	9	12	13	11	12	21	16	17	17	151
Avaria_em_Equipamento/Patrimônio_(O_intencional)	4	12	10	19	13	7	15	15	17	17	20	149
Tentativa_de_suicídio	10	7	6	10	10	12	6	11	13	29	33	147
Achado_de_cadáver	8	16	13	21	12	6	9	17	10	18	14	144
Achado_de_documentos	7	10	16	13	5	8	14	7	19	16	27	142
Corte_irregular_de_árvore	14	12	21	11	16	12	10	7	14	13	12	142
Arma_branca	11	12	16	13	24	11	12	10	11	10	8	138
Óbito	16	19	17	10	3	17	13	11	7	12	13	138
Abandono_de_incapaz	14	8	13	11	11	12	15	5	14	15	16	134
Retirada_de_invasão	18	34	9	5	5	7	17	12	8	8	11	134
Disparo_de_arma_por_GM_com_munição_letal						30	32	25	19	14	13	133
Averiguação_(Defesa_Civil)	12	21	15	14	11	8	5	6	10	20	9	131
Cão_preso_inadequadamente	11	24	21	12	24	5	7	7	10	3	5	129
Encaminhamento_à_FAS	40	17	19	10	7	7	3	1	2	7	14	127
Tentativa_de_agressão	8	10	7	12	7	14	19	10	5	6	14	112
Balão	5	10	7	10	22	19	12	2	7	7	9	110
Rixa	12	11	16	14	5	7	12	5	12	6	10	110
Quedas_de_objetos_ou_partes_de_construções	18	15	11	7	9	11	8	7	5	7	11	109
Fuga_de_aluno/interno	6	6	6	6	11	6	15	10	8	9	25	108
Apoio_ao_Patrimônio	2	6	3	3	8	33	20	19	6	2	5	107
Grafitagem_autorizada					13	49	24	10	5	3	3	107
Tentativa_de_homicídio	12	6	10	9	9	15	10	8	6	6	14	105
Estelionato	5	10	13	16	8	11	8	5	8	7	12	103
Obstrução_de_via	2	7	11	6	8	5	10	1	10	24	19	103
Homicídio	9	14	9	12	10	7	10	4	8	7	5	95
Construção_irregular	11	10	5	2	12	3	5	10	10	11	12	91
Encaminhamento_domiciliar	21	15	8	9	7	4	1	5	6	2	11	89
Risco_de_acidente/à_vida_(Defesa_Civil)	12	5	9	4	9	2	5	8	5	15	14	88
Apoio_à_SMEL	2	3	7	5	9	17	12	9	6	8	9	87

Importunação_ofensiva_ao_pudor	4	4	4	1	4	4	12	7	15	15	17	87
Deslizamento	15	13	14	2	5	7	4	3	3	10	6	82
Apoio_à_SMOP	3	7	10	6	12	5	10	8	6	8	4	79
Estupro	3	5	3	7	7	9	12	14	10	3	6	79
Uso_de_cola/thinner/1ilar	10	11	13	16	11	3	3	1	3		5	76
Verificação	3	7	10	4	6	6	5	5	5	9	12	72
Tentativa_de_invasão_de_equipament o	5	6	7	10	6	7	4	7	5	5	7	69
Achado_de_substância ilícita	1	1	3	3	6	9	12	3	5	7	16	66
Seqüestro_e_cárcere_privado	3	6	8	5	6	7	9	6	8	3	4	65
Escrito_ou_objeto_obsceno_(panfletos_pornográficos)	40	6	11	2	4					1		64
Omissão_de_Cautela_na_Guarda_de_anima is	4	8	5	10	7	10	4	4		4	6	62
Comércio_ambulante	6	7	3	4	5	8	2	4	9	7	6	61
Constrangimento_ilegal	3	7	5	6	2	5	8	6	4	8	6	60
Sem_uso_de_qualquer_instrumento					1	5	9	7	13	9	15	59
Tentativa_de_estupro	3	2	6	5	10	9	5	5	3	3	8	59
Apoio_à_SMAB	1		3	11	10	1	6	6	5	5	10	58
Mendigar, _por ociosidade ou cupidez	2	9	13	8	5	9	4	2	2	1	3	58
Queda_de_fios_de_energia	4	6	7	8	11	2		2	2	8	8	58
Receptação		3	2	3	7	10	7	4	6	5	10	57
Risco_de_queda_de_árvore	9	9	8	11	7	3		1		5	3	56
Vadiagem	17	4	5	9	8	4	7			1	1	56
1ulacro_de_arma_de_fogo	3	8	6	6	5	5	5	4	6	2	5	55
Moeda_falsa	9	1	2	5	1	9	6	9	2	5	6	55
Proteção_ao_patrimônio	3	5	5	1	3	4	4	15	3	5	6	54
Apoio_à_Vigilância_sanitária	6	3	8	6	6	3	5	1	2	12	1	53
Maus_tratos_à_pessoas	5	5	3	7	7	6	6	2	2	2	8	53
Queda_de_galho	8	3	3	7	2	4	4	3	5	7	7	53
Extorsão	2	1	2	3	13	3	13	6	2	2	4	51
Ataque_de_insetos	21	4	3	2	4		1	1	7	2	5	50
Injúria	2	1	2	2	6		8	7	6	6	10	50
Jogo_de_azar	2		4	6	6	3	4	9	4	3	6	47
Falta_de_energia_elétrica	14	11	4		2	4	3	4	3	1		46
Material_abandonado		3	5	7	5	6	9	1	2	3	5	46
Utilizar_carrinho_de_rolimã_em_local _que traga risco	5	5	15	5	7		1	1	3		2	44
Abuso_de_incapazes	2	5	1	2	6	3	5	5	3	4	6	42
Fiscalizações_e_Orientações	2		3	4	5	5	7	7	1	2	6	42
Menor_gazeando_aula	8			5	8	5		3	5	2	6	42
Calote	6	1	5	4	5	4	3		3	3	7	41
Comércio_ilegal_de_aves/animais					18	8	12		1			39
Devolução_de_coisa_achada	2	1	1	5	8	3	1		3	5	10	39
Roubo,_furto,_extravio, _recuperação, _apreensão_de_armas_de_fogo.		5	2	4	4	5	2	1	7	2	7	39
Aliciamento_de_menor	2	4	3	1	2	5	5	3	5	3	3	36

Instituição_financeira	4	4	2	6	5	3	4	1	3		3	35
Apropriação_indébita	1	4	1	3	3	3		5	4	2	8	34
Suicídio	4	2	1	4	3	2	4		5	9		34
Encaminhamento_ao_Conselho_tutelar	3	1	3	3	4	1	7	3	1	5	2	33
Infiltração	8	9	2	1	2	2		2	2	4	1	33
Porte_de_artefato_explosivo	4	1	2	4	7	4	4	1	1	1	4	33
Falsidade_ideológica_(Falsa_Identidade)	3	1	7	3	6	3	4	2	2	1		32
Criador_ilegal_de_aves/animais					13	6	7	1	2			29
Porte_de_cola/thinner/1ilar	10	2	2	3	6	1		2	1		2	29
Apoio_à_Polícia_rodoviária_Federal				1	4		3	11	5	3	1	28
Queda_de_poste	10	1	4	1	3	2	2		1	1	3	28
Apoio_à_COSEDI	1	2	7		2	1			5	4	5	27
Menores_abordando_transeuntes		8	1	1	3	5	3	1	3	1	1	27
Tentativa_de_invasão_de_domicílio_(residência)	1	3		2	3	5	4	1	3	4	1	27
Bicicleta											26	26
Explosão	1	2	5	1	4	3	3	1	2	1	3	26
Por_outras_objetos_ponteados		4	2	3		4	3	3		3	4	26
Dejetos_gerados_por_animais_de_estimação	10	1	4	6	1		3					25
Denúncia_de_bomba	4		3	7		2	6	1			2	25
Resistência		2	2	2		1	1	2		5	10	25
Antecedentes_criminais_-_Verificação							8	3	4	1	8	24
Afogamento	3	2	3		3	4			5		3	23
Apoio_à_SMAD	1	1		3	1	3	1	4	4	3	2	23
Extravio,_sonegação_ou_inutilização_de_livro_ou_doc.	1	3	4	1		1		1		3	7	21
Veículo_conduzido_por_menor	2	1	2	1	4	2	3	2		1	3	21
Porte_de_bebida_alcoólica	2	6	4	4	1	1		1	1			20
Achado_de_munição		2	3	8	1	1		1		1	2	19
Atentado_violento_ao_pudor	7	3	9									19
Impedimento_ou_perturbação_de_cerimônia_funerária		1			3	1	1	4	1	1	6	18
Apoio_à_Polícia_Federal	1				7	2	2	2	1	2		17
Racismo			4	1	1	1	4	1	2	2	1	17
Bueiro_aberto/sem_tampa	3	4		1	1	3		1	1		2	16
Oferecer_bebida_alcoólica_a_menor_de_18_anos	1	2	2	3	3	2		3				16
Risco_de_explosão	1	1	3	3		1	1	1		3	2	16
Achado_de_arma	1		1	2		2	4	2	1	2		15
Achado_de_bomba		2	1	2	2		1	3		1	2	14
Calúnia	3			2		1	1	1	1	1	3	13
Tentativa_de_sequestro			2	1	2	1	1		3	1	2	13
Usar_de_uniforme,_ou_distintivo_de_função_pública_que_o_exerce	1				1	2	4	1	2	2		13
Encaminhamento_à_DP_(sem_flagrant_e)	12											12
Exploração_de_menores		1	2	2	1	1	2	1		1	1	12





Corrupção_ativa		1										1
Defesa_Civil		1										1
Falsificação_de_documento_publico										1		1
Gazuas/chaves_falsas		1										1
Liberação_de_pessoa_presa/apreendi da_por_recusa_no_recebimento_pela _DP										1		1
Óbito_(Defesa_Civil)						1						1
Polícia_Civil										1		1
Prostituição	1											1
Quadrilha_ou_bando										1		1
Rompimento_de_Barragem									1			1
Rufianismo					1							1
Vender_bebida_alcoólica_a_menor_d e_18_anos								1				1
(vazio)												
<b>Total Geral</b>	<b>24.822</b>	<b>22.454</b>	<b>21.388</b>	<b>19.713</b>	<b>24.403</b>	<b>25.250</b>	<b>25.494</b>	<b>20.587</b>	<b>21.534</b>	<b>23.401</b>	<b>32.451</b>	<b>261.497</b>

## APÊNDICE S – SCRIPT DO R STUDIO AGRUPAMENTO DA BASE PELO K-MEANS

```
##### Carregando os dados #####
dados <- read.csv2('dadosGuarda.csv')

#Esta função data.table possibilita mais funções que a função data.frame (padrão do R)#
dados2 <- data.table(dados)

##### Agrupamento com k-means #####
grupos <- fastDummies::dummy_cols(dados2$NATUREZA1_DESCRICAO)
grupos <- grupos[,-1] # remove a primeira coluna de texto

d <- dist(t(grupos)) # t transpõe a matriz

km <- kmeans(d, 5) # defini o numero de grupos
table(km$cluster)

View(km$cluster)
write.csv2(km$cluster, "kmeans.csv")
```

## APÊNDICE T – SCRIPT DO R STUDIO ALGORITMO ÁRVORE DE DECISÃO

```

1 #####Instalando pacotes #####
2 install.packages("data.table")
3 install.packages("rpart")
4 install.packages("rpart.plot")
5 install.packages("prp")
6
7 ##### Carregando pacotes #####
8 library(data.table)
9 library(rpart)
10 library(rpart.plot)
11 library(prp)
12 #carrega os arquivos CSV
13 dadosAgrManual <- read.csv2('Guarda_manual.csv')
14 dadosAgrKmeans <- read.csv2('Guarda_Kmeans.csv')
15 |
16 #Esta função data.table possibilita mais funções que a função data.frame (padrão do R)
17 Manual <- data.table(dadosAgrManual)#Dados com o Agrupamento Manual
18 K_means <- data.table(dadosAgrKmeans)#Dados com o agrupamento do K-Means
19
20 #Algoritmo da Arvore de Decisão
21 Arvore <- rpart(CODIGODESCRICAO ~ REGIONAL_FATO_NOME + ATENDIMENTO_ANO + OCORRENCIA_DIA_SEMANA +
22 OCORRENCIA_HORA + OCORRENCIA_MES + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL + ORIGEM_CHAMADO_DESCRICAO ,
23 data = Manual, method = 'class')
24
25 Arvore <- rpart(CODIGODESCRICAO ~ REGIONAL_FATO_NOME + ATENDIMENTO_ANO + OCORRENCIA_DIA_SEMANA +
26 OCORRENCIA_HORA + OCORRENCIA_MES + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL + ORIGEM_CHAMADO_DESCRICAO ,
27 data = K_means, method = 'class')
28
29 #Vizualização da arvore
30 Arvore #visualiza os resultados no console
31 prp(Arvore,type=4, extra=102) #visualiza a arvore com a probabilidade e rótulos de divisão abaixo dos rótulos dos nós
32 rpart.plot(Arvore,type=4, extra=102)#visualiza a arvore com simplificada
33
34
35 #Matriz confusão base de dados Manual
36 matriz <- predict(Arvore, Manual, type = "class")
37 table( Manual$CODIGODESCRICAO, matriz)
38
39 #Matriz confusão base de dados K_means
40 matriz <- predict(Arvore, K_means, type = "class")
41 table( K_means$CODIGODESCRICAO, matriz)

```

## APÊNDICE U – SCRIPT DO R STUDIO ALGORITMO NAÏVE BAYES

```

1 ▾ ##### Instalando pacotes #####
2 install.packages("data.table")
3 install.packages("dplyr")
4 install.packages("psych")
5 install.packages("caret")
6 install.packages("stringr")
7 install.packages("naivebayes")
8 install.packages("e1071")
9 ▾ ##### Carregando pacotes #####
10 library(data.table)
11 library(dplyr)
12 library(psych)
13 library(caret)
14 library(stringr)
15 library(naivebayes)
16 library(e1071)
17 #carrega os arquivos CSV
18 dadosAgrManual <- read.csv2('Guarda_manual.csv')
19 dadosAgrKmeans <- read.csv2('Guarda_Kmeans.csv')
20
21 #Esta função data.table possibilita mais funções que a função data.frame (padrão do R)
22 Manual <- data.table(dadosAgrManual)#Dados com o Agrupamento Manual
23 K_means <- data.table(dadosAgrKmeans)#Dados com o agrupamento do K-Means
24
25 #Algoritmo da Naive Bayes
26 Bayes <- naiveBayes(CODIGODESCRICA0 ~ ATENDIMENTO_ANO + OCORRENCIA_MES + OCORRENCIA_DIA_SEMANA
27 + OCORRENCIA_HORA + REGIONAL_FATO_NOME + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL
28 + ORIGEM_CHAMADO_DESCRICA0 ,data = Manual)
29
30 Bayes <- naiveBayes(CODIGODESCRICA0 ~ ATENDIMENTO_ANO + OCORRENCIA_MES + OCORRENCIA_DIA_SEMANA
31 + OCORRENCIA_HORA + REGIONAL_FATO_NOME + FLAG_FLAGRANTE + NATUREZASIM_DEFESA_CIVIL
32 + ORIGEM_CHAMADO_DESCRICA0 ,data = K_means)
33 #Vizualização do resultado
34 Bayes
35
36 #Matriz confusão
37 matriz1 <- predict(Bayes, Manual)
38 table( Manual$CODIGODESCRICA0, matriz1)
39
40 #Matriz confusão
41 matriz2 <- predict(Bayes, K_means)
42 table( K_means$CODIGODESCRICA0, matriz2)

```